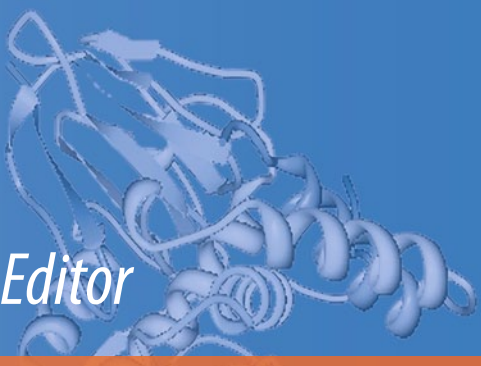
Subcellular Biochemistry 68



Mauricio G. Mateu *Editor*

Structure and Physics of Viruses

An Integrated Textbook

Springer

Structure and Physics of Viruses

SUBCELLULAR BIOCHEMISTRY

For further volumes:
http://www.springer.com/series/6515

Mauricio G. Mateu
Editor

Structure and Physics of Viruses

An Integrated Textbook

Springer

Editor
Mauricio G. Mateu
Centro de Biología Molecular
"Severo Ochoa" (CSIC-UAM)
and Department of Molecular Biology
of the Universidad Autónoma de Madrid
Madrid, Spain

ISSN 0306-0225
ISBN 978-94-007-6551-1 ISBN 978-94-007-6552-8 (eBook)
DOI 10.1007/978-94-007-6552-8
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013939581

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The great challenges of modern science are increasingly requiring the combined efforts of different branches of knowledge. A remarkable example is our long-standing battle to fight and understand viruses, these tiny pathogens that are a permanent companion and threat for living beings and humankind. Viruses have been traditionally a subject of study for biologists, but are increasingly fascinating physicists and chemists alike, dazzled by their clever operation and their outstanding potential for promising applications. What makes viruses special is that they are comparatively simple in the biological context, yet extremely effective in replicating themselves in all kinds of hosts. Viruses have found optimal solutions to survive in very harsh conditions and are teaching us important lessons on how to work efficiently at the nanoscale, making the most of their very limited resources.

A large part of our present knowledge of viruses has been facilitated by techniques, theories, and methods developed in physics and chemistry. In biology, structure is the key to understand function and physical techniques like X-ray diffraction or cryo-electron microscopy have opened the door to unveil structural details of viruses with nearly atomic resolution. In addition, function ultimately involves a set of processes that cannot escape physical laws and that, in fact, very often make a profitable use of them. In recent years, physics and chemistry are not only helping in providing new experimental tools to investigate viruses but are also increasingly contributing to achieve a qualitative and quantitative comprehension of the processes involved. This close collaboration between physics, chemistry, and biology has led to great advances and has boosted our present understanding of viruses.

But interdisciplinarity has its toll. It is often very hard for a scientist trained in one of the traditional sciences to navigate between two worlds, overcoming the different points of view and the methodological and terminological barriers between disciplines. In this context, a book like this constitutes an invaluable resource to cover this gap for students and scientists entering the fascinating world of viruses. The book gathers a complete selection of chapters written by leading experts in the field. It has been written primarily with a nonexpert audience of students and researchers in mind; however, it has been aimed also at providing

practitioners in one area of structural or physical virology with updated overviews of most other areas of these broad scientific disciplines. It covers distinct but complementary aspects of the study of viruses that go from techniques required for their characterization to their potential applications, including its structure and functioning.

This effort has been partially possible thanks to the *Spanish Network on the Biophysics of Viruses* (BioFiViNet), supported by the Spanish Ministry of Science. This novel initiative is aiming at coordinating the efforts of the Spanish physicists, chemists, and biologists interested in the interdisciplinary study of viruses and has been the initial spark to agglutinate most of the experts that ended up being authors of this book. It has also promoted the added value of collaborative work between apparently widely different disciplines, thanks to the support and enthusiasm of all its members.

But, without doubt, the real artificer of this great work is its editor Mauricio G. Mateu, who has meticulously planned, coordinated, and assisted the efforts of all authors to accomplish this interdisciplinary and formative spirit. In the name of all authors, I would like to thank Mauricio for his relentless effort and countless hours of dedication to this book. I am sure that future physical and structural virologists will also appreciate his efforts, and that this book will help them to prepare for the myriad wonderful lessons that viruses still have to teach us.

Barcelona, Spain
November 2012

David Reguera

Preface

Structural Virology is today an all-important discipline that permeates most other virological disciplines. The application of physical and physicochemical techniques has led to the determination of the high-resolution molecular structures of many viruses. The interplay of this approach with (bio)chemical and biological approaches has allowed in many cases the elucidation of the structural basis of viral function in unprecedented detail. In addition, in the last years, theoretical and experimental physicists have begun to tackle a fundamental physics-based approach to study different aspects of the architecture, self-assembly, and material properties of virus particles. A new term, *Physical Virology*, has recently been coined to encompass these and related studies. This approach is slowly beginning to merge with long-standing structural virology approaches to provide a renewed and richer view on viruses.

Since I became interested in virus structure–function relationships over 25 years ago, I have had the privilege of collaborating not only with virologists but also with molecular and structural biologists, organic and physical chemists, and theoretical and condensed-matter physicists interested in viruses. Since 8 years ago, my group is involved in a multidisciplinary network of physicists, chemists, and biologists to study nano-objects, coordinated first by Prof. Fernando Flores and later by Prof. Julio Gómez-Herrero and funded by the Comunidad de Madrid regional government. As collaborations progressed, I became aware of the difficulties we sometimes faced in finding common scientific goals and interdigitating research pathways instead of merely crossing scientific trajectories. Like many others in similar situations I happily found that, once we began to understand each other's scientific language and aims, we were able to tackle together fruitful combined studies on viruses. This experience and talks with colleagues and students convinced me, about 3 years ago, on the usefulness of an interdisciplinary textbook in which the rapidly expanding fields of structural and physical virology were dealt with in an integrated way. The opportunity to realize such a project came from a kind invitation in early 2011 by Dr. Thijs van Vlijmen at Springer.

At about the same time Prof. David Reguera had the idea of creating in Spain a national network on the biophysics of viruses (*BioFiViNet*) to reinforce previous

bonds and promote further interactions between physicists, chemists, and biologists doing research on virus biophysics. Also, a Master in Virology in which many structural virologists were invited to participate had recently been organized under the auspices of the Spanish Society for Virology presided by Prof. Esteban Domingo. I decided it was high time to try and convince some of my colleagues at BioFiViNet and the Spanish Society for Virology, particularly those who had been or are presently engaged in collaborative research among us, to jointly write an advanced textbook on the structure and physics of viruses. I contacted 20 close colleagues (including physicists, chemists, and biologists) from 12 Spanish institutes or university departments. All of them are active senior researchers and internationally recognized experts at the cutting edge of their research areas within structural or physical virology. They all accepted, and some recruited other close colleagues to coauthor their chapters. Such general closeness among book authors has allowed intense mutual feedback during the organization and writing of the book through frequent e-mails, phone calls, and meetings of the editor with the authors, and also between authors of different chapters. Every chapter has been critically read by the editor and (in nearly all cases) by at least one author of a different chapter as well. As a result, each chapter has been written and revised considering the detailed contents of the other chapters, and relating to them as much as possible and practical. However, some minor overlaps between chapters have been kept to facilitate the understanding of each chapter subject without a need to read other chapters in the book. We feel these minor overlaps may also help in connecting the different chapter's contents.

After an infectious virus particle (a virion) targets a host cell, it loses its integrity and releases its nucleic acid genome inside the cell. In the period of time that goes from that point until the progeny viruses are formed, the virus as a discrete physical entity ceases to exist, but the viral genetic instructions inside the infected cell may subvert the cell metabolism. Eventually, many copies of the viral genome and of the viral proteins required to form a new virus particle will be made in the cell, and a number of progeny virions will be assembled from them, closing an infectious cycle (or viral "life" cycle). The structural biology of viral metabolic processes such as replication, recombination, integration, transcription, or translation and their spatial and temporal regulation, and the virus-induced alteration of cellular components and reactions, are outside the scope of this book. Structure-based aspects regarding these processes are integrated in many excellent Molecular Virology books, a few of which are referred to at the end of the introductory chapter in this book (Chap. 1).

The present book has been focused instead on the "other half" of the viral cycle, which is generally less well known. Specifically, this book contemplates the structure, dynamics, and physics of virus particles: From the moment they come into existence by self-assembly from viral components produced in the infected cell, through their extracellular stage, until they recognize and infect a new host cell and cease to exist by losing their physical integrity to start a new infectious cycle. (Bio)physical techniques used to study the structure of virus particles and components and some applications of structure-based studies of viruses are also contemplated.

This book is aimed first at M.Sc. students, Ph.D. students, and postdoctoral researchers with a university degree in biology, chemistry or physics or related sciences who share an interest or are actually working on viruses. We have aimed also at providing an updated account of many important concepts, techniques, studies, and applications in structural and physical virology for established scientists working on viruses, irrespective of their physical, chemical, or biological background and their field of expertise. We have *not* attempted to provide a collection of *for-experts-only* reviews focused mainly on the latest research in specific topics; we have *not* generally assumed that the reader knows all of the jargon and all but the most recent and advanced results in each topic dealt with in this book. In short, we have attempted to write a book basic enough to be useful to M.Sc. and Ph.D. students, as well as advanced and current enough to be useful to senior scientists.

Inevitably, some compromises had to be made. Because of space limitations, not every possible topic has been contemplated; however, we believe most of the important general aspects of the structure and physics of virus particles as they are known have been covered. Space limitations have also prevented the authors of the different chapters to include explanations of some elementary or general concepts or terms. However, we believe the most important aspects in each chapter will be clearly understandable by those with a B.Sc.-level knowledge of physics, chemistry, biology or related scientific areas. Quick consultation to *Wikipedia* or other general sources may solve an occasional doubt on a specific term by a reader coming from a different area of knowledge. In any case, several teaching aids have been implemented in the book to facilitate the reading and understanding of each chapter by a newcomer to the field; these aids include:

An introductory chapter. Chapter 1 includes a brief general introduction to viruses and their structure, some basic concepts and terms in molecular and structural virology, and general descriptions of different steps in the virus cycle. The chapter is also intended as a guide to help the reader integrate in a general picture the topics treated in each monographic chapter (Chaps. 2–22).

A similar outline in different chapters in each part of the book. Each Part II chapter dealing with a specific technique includes sections on the principles of the technique, on relevant examples of contributions of the technique for understanding viruses, and on technical perspectives. Each Part III chapter dealing with a stage in the viral cycle includes sections that connect the stage described with stages described in other chapters.

Frequent cross-references between chapters. They may be useful to find in other chapters additional information on certain subjects and/or to connect particular aspects treated in different chapters.

Basic systematic information on viruses species and families mentioned in the book. This information is intended mainly to help the reader navigate among the multitude of virus names that will inevitably appear in the different chapters of this book (or of any other virology book). Table 1.1 includes (nearly) all virus species, families, and orders mentioned in this book, with indication of host and viral genome types. Figure. 1.3 gives a scheme of the general virion structure for some

of the most important families of animal viruses. *Specific indexes* at the end of the book include (nearly) all virus species and families mentioned, each with reference to pages where some information regarding the species or family is mentioned.

A list of references focused largely on review articles. The list of references at the end of each chapter has purposefully been kept relatively short in order not to overwhelm the nonexpert reader; reviews have been included wherever possible, and the advanced reader is referred in many cases to those reviews for the original references that have not been included in the book. We apologize to the many authors whose work, however important, could not be directly cited.

A section on further reading. In each chapter, a short list under the subheading *further reading* includes references to a few books and/or reviews. These may be particularly useful either for learning or refreshing basic principles, or for more detailed/advanced information on the subject.

This book is organized into four parts. Although the chapters are self-contained and may be understood individually, all chapters, especially those in each part (I–IV), are inter-related and they have been designed with the input of the editor and every author to tell a complete story.

Part I, *The viral machine*, includes an introductory chapter to the rest of the book (Chap. 1) and another chapter on the fundamental composition and basic architecture of virus particles (Chap. 2). This knowledge is essential for understanding many subjects treated in the following chapters.

Part II, *Determination of the structure and physical properties of viruses*, contemplates most of the major experimental techniques in structural and physical virology. These include different electron microscopy techniques, with emphasis in cryo-electron microscopy and tomography (Chap. 3); X-ray crystallography (Chap. 4); nuclear magnetic resonance spectroscopy (Chap. 5); other spectroscopic techniques (circular dichroism and fluorescence) and mass spectrometry (Chap. 6); the combination of the above and other structural biology methods (Chap. 7); and single-molecule techniques, including atomic force microscopy (Chap. 8) and optical tweezers (Chap. 9). In all cases, the techniques are described not in broad, general terms (as can be found in general technical books) but in the ways they are specifically applied to study virus particles and their components.

Part III, *Structural foundations of virus properties and functions*, deals with the different stages in the viral cycle in which virus particles and/or their structural components are involved. Confronting the *chicken and egg* problem, we decided to start the endless viral cycle when a viral particle comes into existence by self-assembly and proceed until the viral particle ceases to exist by losing its integrity and releasing its nucleic acid in a host cell. As previously noted, the metabolism of the viral nucleic acid and other viral components until new viral particles are formed (thus closing the viral cycle) does not involve virus particles and is, thus, out of the scope of this book. Chapters 10–17 consider each viral cycle stage in which virus particles are involved. In most of these chapters, the authors have focused in studies on some model viruses that constitute paradigms to understand the structural bases of virus function at that stage in the cycle. In addition, some general conclusions are extracted from the cases described and other studies.

Chapter 10 deals with experimental studies on the basic assembly of structurally simple viruses; both structural aspects and the cellular environment where assembly occurs are contemplated. Chapter 11 builds on what has been described in Chaps. 2 and 10 on relatively simple viruses, to describe important specific aspects of the structure and assembly of more complex viruses, including their scaffold-mediated assembly. Chapter 12 focuses on the different ways in which the viral nucleic acid is packaged inside the virus particle. Chapter 13 deals with the process of maturation of viral particles to become infectious virions. Self-assembly or assisted assembly of a virus shell (capsid), packaging of the nucleic acid, and virus particle maturation are not clear-cut steps and they frequently overlap in space and time; thus, particularly frequent cross-references between Chaps. 10–13 have been included. Chapter 14 recollects a wide range of structural, molecular, and cell biology techniques as they are applied to understand the complete virus morphogenetic process as it occurs *in vivo*; this chapter also describes several important observations obtained on this still little-known process.

No specific chapter in Part III is devoted to the properties of the extracellular mature virions because they are sufficiently covered in Chap. 1 and/or in other chapters dealing with processes in which those properties come into play. Virion conformational stability and dynamics are revised in Chap. 1 in the context of the complete viral cycle and are dealt with also in other chapters as related with some particular stages in the infectious cycle. Several aspects of the recognition of extracellular virus particles by the immune system, mechanisms of virus recognition and neutralization by antibodies and virus escape from antibody recognition are briefly revised in Chaps. 21 and 1, respectively.

Chapters 15–17 describe the complex process of virus entry into cells and viral nucleic acid transfer. Chapters 15 and 16 deal with the specific recognition of host cells by animal viruses through interaction with receptor molecules, the entry of nonenveloped or enveloped viruses into these cells by different mechanisms, and the release (uncoating) of the viral genome inside the cell. Chapter 17 describes the mechanisms bacteriophage viruses follow for bacterial cell recognition and transfer into the cell of their genome.

The last chapters in Part III (Chaps. 18 and 19) deal with salient studies on the fundamental physics of viruses. Chapter 18 describes recent experimental studies on the mechanical properties of virus particles and their possible impact on virus biology. Chapter 19 contemplates theoretical physics-based studies on the more fundamental aspects of virus architecture, material properties, assembly and entry into host cells, all of them highly relevant subjects about which some experimental studies have been described in previous chapters (Chaps. 2 and 10–18). Readers of any of those chapters are strongly encouraged to read also Chap. 19 to acquire a more complete understanding of the structure and properties and functions of virions, and the synergy between experimental and theoretical, structural and physical virology studies.

Part IV, ***Applied structural and physical virology***, includes some important current or potential developments and applications of the structural and physical knowledge being acquired on viral particles. We chose three major applied areas.

Chapter 20 contemplates the general design and structural basis of action of antivirals. In this chapter, antivirals against molecules other than virus particles or their components are contemplated first, because the design principles and structural basis of action do not largely depend on the target, and some of those antivirals have been well-studied and proved remarkably successful; the chapter then continues by considering novel approaches based on the inhibition or misdirection of virus entry or morphogenesis. Chapter 21 describes novel approaches to vaccines based on virus-like particles or chimeric virions. Chapter 22 contemplates the chemical and genetic manipulation and the use of viral particles for applications in the rapidly expanding nanotechnology field, from biomedicine to electronics.

I would like to gratefully acknowledge the authors of this book for their enthusiasm, time, effort, and patience devoted not only to write their chapters but to meet and discuss contents, presentation, and improvements; Prof. David Reguera for writing the foreword and for critically reading many book chapters; Miguel Angel Fuertes for most helpful assistance with a number of tasks; and José A. Pérez for formatting some figures. The authors of this book wish to collectively express here our gratitude to those other book authors who critically read our chapter's manuscripts. We are indebted to Dr. Thijs van Vlijmen at Springer for his support and interest in this book project, and to him, Ms. Sara Germans at Springer and Mr. Ibrahim Mohamed Asif at SPi Technologies for their invaluable help and patience during the writing, edition and production of this book.

Madrid, Spain
November 2012

Mauricio G. Mateu

Contents

Contributors

Nicola G.A. Abrescia Structural Biology Unit, CICbioGUNE, CIBERehd, Derio, Spain

Ikerbasque, Basque Foundation for Science, Bilbao, Spain

José M. Almendral Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), and Department of Molecular Biology, Universidad Autónoma de Madrid, Madrid, Spain

José María Alonso CIC nanoGUNE, San Sebastián, Spain

J. Ricardo Arias-Gonzalez Instituto Madrileño de Estudios Avanzados en Nanociencia (IMDEA Nanociencia), Madrid, Spain

Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

CNB-CSIC-IMDEA Nanociencia associated unit "Unidad de Nanobiotecnología", Madrid, Spain

Daniel Badía-Martínez Structural Biology Unit, CICbioGUNE, CIBERehd, Derio, Spain

Juan Bárcena Centro de Investigación en Sanidad Animal (INIA), Madrid, Spain

Alexander M. Bittner CIC nanoGUNE, San Sebastián, Spain

Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Esther Blanco Centro de Investigación en Sanidad Animal (INIA), Madrid, Spain

José L. Carrascosa Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

José M. Casasnovas Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

José R. Castón Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Ana Cuervo Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

María I. Daudén Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Laura R. Delgui Laboratorio de Biología Celular y Molecular, Instituto de Histología y Embriología Mendoza (IHEM), Facultad de Ciencias Médicas, Universidad Nacional de Cuyo-CONICET, Mendoza, Argentina

Instituto de Ciencias Básicas, Universidad Nacional de Cuyo, Mendoza, Argentina

Isabel Fernández de Castro Cell Structure Laboratory, Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Pedro J. de Pablo Department of Physics of the Condensed Matter, C-III, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

Ignacio Fita Institut de Biología Molecular de Barcelona (CSIC), and IRB Barcelona, Parc Científic de Barcelona, Barcelona, Spain

Federico Gago Department of Pharmacology, Universidad de Alcalá, Madrid, Spain

Carmela Garcia-Doval Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Damià Garriga Department of Molecular and Cell Biology, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Marcin Ł. Górzny CIC nanoGUNE, San Sebastián, Spain

Antoni Luque Department of Fundamental Physics, Universitat de Barcelona, Barcelona, Spain

Department of Chemistry, New York University, New York, NY, USA

Vicente Más Biología Viral, Centro Nacional de Microbiología and CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III Majadahonda, Madrid, Spain

Mauricio G. Mateu Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), and Department of Molecular Biology, Universidad Autónoma de Madrid, Madrid, Spain

José A. Melero Biología Viral, Centro Nacional de Microbiología and CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III Majadahonda, Madrid, Spain

Luis Menéndez-Arias Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Madrid, Spain

José L. Neira Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Alicante, Spain

Instituto de Biocomputación y Física de Sistemas Complejos, Zaragoza, Spain

Hanna M. Oksanen Institute of Biotechnology and Department of Biosciences, Viikki Biocenter, University of Helsinki, Helsinki, Finland

David Reguera Department of Fundamental Physics, Universitat de Barcelona, Barcelona, Spain

Cristina Risco Cell Structure Laboratory, Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

José F. Rodríguez Department of Molecular and Cell Biology, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Carmen San Martín Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

David I. Stuart Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, UK

Diamond Light Source Ltd, Diamond House, Didcot, UK

Mark J. van Raaij Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), Madrid, Spain

Nuria Verdaguer Institut de Biología Molecular de Barcelona (CSIC), Parc Científic de Barcelona, Barcelona, Spain

Christina Wege Institute of Biology, University of Stuttgart, Stuttgart, Germany

Part I
The Viral Machine

Chapter 1
Introduction: The Structural Basis of Virus Function

Mauricio G. Mateu

Abstract Viruses may be regarded as dynamic nucleoprotein assemblies capable of assisted multiplication within cells, and of propagation between cells and organisms. Infectious virus particles (virions) assembled in a host cell are dynamic, generally metastable particles: They are robust enough to protect the viral genome outside the cell, but are also poised to undergo structural changes and execute mechanochemical actions required for infection of other cells. This chapter provides an introduction to the structural and physical biology of viruses by including: (i) an elementary overview on virions and the structural basis of virus function; (ii) a concise summary on basic techniques used in structural or physical virology; (iii) brief structure-based general descriptions of the different stages in the virus cycle, especially those in which virions and/or their components are involved. These contents may facilitate a better understanding of the specialized subjects treated in the rest of the book. This chapter is also intended as a "road map" to help interconnect and integrate in a single picture the different topics described in depth in the 21 monographic chapters in this book.

Keywords Virus • Virus cycle • Infection • Capsid • Viral genome • Capsid subunits • Capsid building blocks • Oligomerization • Self-assembly • Assisted assembly • Assembly intermediates • Scaffolding proteins • Conformational stability • Conformational dynamics • Nucleic acid packaging • Capsid-nucleic acid condensation • Virus maturation • Virus stability and dynamics • Virus-antibody recognition • Virus-receptor recognition • Virus entry • Fusion • Uncoating • Antivirals • Vaccines • Biotechnology • Nanotechnology

M.G. Mateu (✉)
Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM)
and Department of Molecular Biology of the Universidad Autónoma
de Madrid, c / Nicolás Cabrera 1, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: mgarcia@cbm.uam.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*,
Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_1,

AAV	Adeno-associated virus
AFM	Atomic force microscopy
CBB	Capsid building block
CP	Capsid protein
cryo-EM	Cryo-electron microscopy
DNA	Deoxyribonucleic acid
ds	Double-stranded
EMDB	Electron Microscopy Data Bank
Fab	Antigen-binding antibody fragment
FMDV	Foot-and-mouth disease virus
HIV-1	Human immunodeficiency virus type 1
HRV	Human rhinovirus
HSV-1	Herpes simplex virus type 1
ICTV	International Committee on Taxonomy of Viruses
MD	Molecular dynamics
mRNA	Messenger RNA
MS	Mass spectrometry
MVM	Minute virus of mice
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
PV	Poliovirus
RNA	Ribonucleic acid
RT	Reverse transcriptase
SAXS	Small-angle X-ray scattering
ss	Single-stranded
TMV	Tobacco mosaic virus
VLP	Virus-like particle

1.1 The Structure and Physics of Viruses

Viruses are biological entities capable of assisted multiplication within cells and of propagation between cells and organisms. Virus multiplication and propagation is generally a cyclic process: An infectious viral particle (*virion*) introduces its genome into a host cell, new virions are formed in the cell and released, and these in turn may infect other host cells. This cycle of infection is often called the *virus* "*life*" *cycle*. There has been a largely philosophical debate on whether viruses are alive or not. We use the term *virus life cycle* as a synonym of *infectious cycle*; we are not making the statement that viruses are "living" organisms. In this book viruses are contemplated as macromolecular complexes that, through biological evolution, came into existence and were endowed with the capacity to make copies of themselves by using the genetic instructions they enclose and the host cell machinery.

Because of the effects many viral infections cause on living beings, viruses are frequently considered only as pathogens causing disease and human suffering, economic losses and social problems. However, since the times of the "Phage Group" and the advent of molecular virology more than half a century ago, scientists coming from different areas have become increasingly aware that viruses also provide outstanding, relatively simple models to explore biomolecular structure-function relationships using a combination of physical, chemical and biological approaches. The knowledge thus acquired has been decisive not only to combat viral disease, but also in the quest to understand in physico-chemical terms the molecular machinery of life.

Specific reasons for studying the structure, dynamics and physical and (bio) chemical properties of virus particles include the following:

(i) Virus particles constitute excellent models to understand and learn to manipulate molecular self-assembly.
(ii) Virus particles are paradigms to understand structure-function relationships in biomacromolecular assemblies and biological machines.
(iii) A profound knowledge of virus structure, dynamics and properties is essential to understand the life cycles of viruses.
(iv) Virus particles, their components and the processes in which they participate provide novel targets for the design of antiviral agents.
(v) Understanding the structural determinants of virus stability, dynamics and function may facilitate the rational manipulation of virus particles to develop new or improved vaccines, gene therapy vectors, and nanoparticles for drug delivery or other biomedical or bio/nanotechnological uses.

This chapter provides an introduction to structural and physical virology and is intended mainly for M.Sc. students, Ph.D. students and postdoctoral researchers in physics, chemistry, biology or related areas who are interested in viruses, but who may be relatively unfamiliar with the subject. It intends also to provide a "road map" to help the reader integrate in a general picture the topics treated in depth in the other chapters in this book. To achieve these aims, the present chapter includes:

(i) Brief explanations on some elementary concepts and terms in molecular, structural and physical virology; a detailed description of the basic architecture of viruses will follow in the accompanying Chap. 2 in Part I of this book.
(ii) Some broad guidelines on the applicability of most of the different techniques described in Part II, Chaps. 3, 4, 5, 6, 7, 8 and 9, and Chaps. 14, 19, to characterize the structure, dynamics and physical properties of virus particles.
(iii) A brief overview of a generic virus cycle and of the major roles in the cycle of virions and their components.
(iv) Brief accounts of general structural concepts and descriptions regarding each of the different stages of the viral cycle in which virions or their components are involved, and of relevant properties of virus particles. These short, elementary accounts may facilitate those readers with little background in molecular and structural virology a fullest and better integrated understanding of the

contents of Part III, Chaps. 10, 11, 12, 13, 14, 15, 16 and 17; each of these chapters deals in detail with structural aspects of one of those stages of the virus cycle.

(v) A schematic overview of novel, physics-based approaches to study virus structure, dynamics and properties, as a brief introduction to detailed accounts of some physical virology methods (Part II, Chaps. 8, 9), studies (Part III, Chaps. 18, 19) and applications (Part IV, Chap. 22).

(vi) A very brief overview of applied studies in structural virology, to put into a general context the particular applications described in detail in Part IV, Chaps. 20, 21 and 22.

1.1.1 Structural Virology

Our knowledge of the molecular structure and function of viruses has grown spectacularly in the last decades, largely because these entities have uniquely and increasingly attracted the interest of biologists, (bio)chemists and physicists alike. Viruses have, thus, been rediscovered as organized complexes of biomolecules that act as minute machines. These nanomachines are continuously being modified and diversified through mutation and biological adaptation. However, they are invariably determined by the laws of physics and chemistry to blindly perform sophisticated mechanochemical actions, including penetration into host cells and self-assembly from their molecular components after these have been replicated in the cell. The application of physical and physicochemical techniques has led to the determination of the high-resolution molecular structures of many viruses; the interplay of this approach with (bio)chemical and biological approaches have allowed in many cases the elucidation of the structural basis of viral function in unprecedented detail. *Structural Virology* has permeated other virological disciplines to provide a molecular view of viruses and their biology. The detailed structural knowledge on viruses and their components has made, and will surely continue to make, decisive contributions in the fight against viral disease.

1.1.2 Physical Virology

In the last years, the advent of nanoscience and nanotechnology, and the increasing awareness on the outstanding features of virus particles as solid-state nanodevices are leading to a renewed look at viruses from the physicist's point of view. Theoretical physicists have begun to tackle at the most fundamental level different aspects of the architecture, self-assembly and physical properties of virus particles. Also, the development of atomic force microscopy (AFM), optical tweezers and other techniques to study individual molecules have opened up new possibilities for the experimental study of the structure, properties and mechanochemical actions of

viruses and their components. A new term, *Physical Virology*, has recently been coined to encompass theoretical and experimental physics-oriented studies of viruses. This novel approach is slowly beginning to merge with long-standing structural virology approaches based on other physical or physico-chemical techniques such as electron microscopy, X-ray crystallography and many others. As a consequence, viruses are currently being investigated for new developments not only in biomedicine and biotechnology but also in nanotechnology, including nanomaterials and nanoelectronics.

1.2 Virions and Their Structural Components

1.2.1 Molecular Composition of Viruses

From a structural point of view virions may be generally regarded as nucleoprotein assemblies. They all include a nucleic acid genome and many copies of one or more proteins. However, virions present remarkable differences in size, shape, molecular composition, structural organization and complexity (Fig. 1.1). Considering its molecular composition only, viruses are generally classified into two large groups, *non-enveloped viruses* and *enveloped viruses*, depending on the absence or presence of an outer lipid layer.

Non-enveloped Viruses

The simplest non-enveloped virions are composed just of a protein shell, or *capsid* (sometimes called *coat*) made of multiple copies of one or more proteins, that contains the viral nucleic acid (Fig. 1.1, left). In less simple non-enveloped virions, the capsid may contain not only the viral genome, but also other proteins and/or other macromolecules, which may be organized in subassemblies. These additional biomolecules or subassemblies can be enclosed in the capsid shell or externally attached to its surface (Fig. 1.1, center) (see Chaps. 2, 10, 11, 17).

Enveloped Viruses

In enveloped virions, the capsid and/or other internal structures are typically surrounded by a lipid bilayer, or *envelope*, in which some proteins are embedded (Fig. 1.1, right); some enveloped virions have a complex multi-layer structure made of organized lipid, protein and/or nucleoprotein layers (see Chaps. 2, 11).

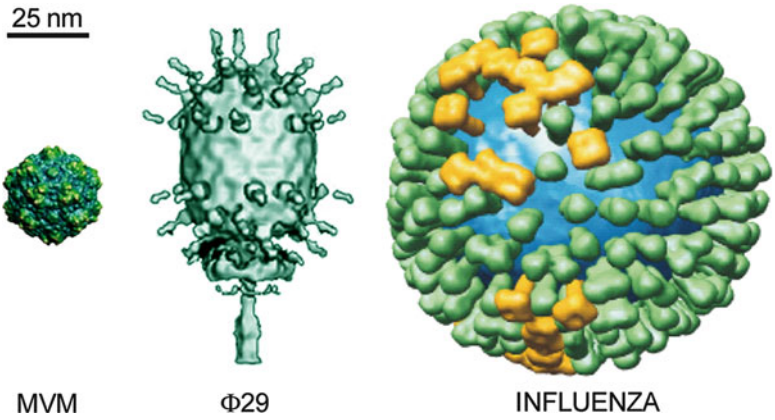


Fig. 1.1 Major types of viruses by molecular composition. *Left* and *center*, non-enveloped viruses; *left*, a very simple virus (the parvovirus MVM); *center*, a complex virus (the tailed bacteriophage ϕ29). *Right*, an enveloped virus (the orthomyxovirus influenza virus). They are reproduced at (approximately) the same scale, indicated by the *horizontal bar* (*top left*). The MVM structural model [Agbandje-McKenna M, Llamas-Saiz AL, Wang F, Tattersall P, Rossmann MG (1998) Structure 6:1369–1381] was obtained from VIPERdb [Carrillo-Tripp M, Sheperd CM, Borelli IA, Sangita V, Lander G, Natarajan P, Johnson JE, Brooks III CL, Reddy VS (2009) Nucleic Acids Res 37:D436–D442]. The models of ϕ29 and influenza virus are respectively reproduced from [Wikoff WR, Johnson JE (1999) Curr Biol 9:R296–R300] and [Harris A, Cardone C, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC (2006) Proc Natl Acad Sci USA 103:19123–19127], with permission. (Figure kindly provided by M.A. Fuertes)

1.2.2 The Virus Capsid

The capsid plays a fundamental role in both the architecture and the biological function of a virus. A virus capsid can be generally described as a hollow symmetric protein oligomer or multimer made from several tens to many hundreds of copies of one or a few different types of folded polypeptides, the *capsid protein* (CP) subunits. Most CPs (or their individual structural domains if formed by more than one domain) can be ascribed to one of a very limited number of protein architectures or *folds* that can assemble into a limited number of quaternary structures (Chap. 2).

In each virus, oligomerization of the CPs during capsid assembly normally leads to a defined type of symmetric quaternary structure. Only very few types of capsid symmetry are frequent. The basic types are *helical* (Fig. 1.2, top) and *icosahedral* (Fig. 1.2, bottom left and right). In some viruses, capsids adopt an elongated (prolate) icosahedral architecture (Fig. 1.1, right). Other capsids, such as those of retroviruses or poxviruses, are made of less simple arrangements of CP subunits (see Chaps. 2, 10, 11). Structural models (density maps) of virions and other virus particles determined by electron microscopy (Chap. 3) are available at the Electron Microscopy Data Bank (EMDB) (http://www.ebi.ac.uk/pdbe/emdb). Atomic coordinates and high-resolution structural models of virions or other virus particles

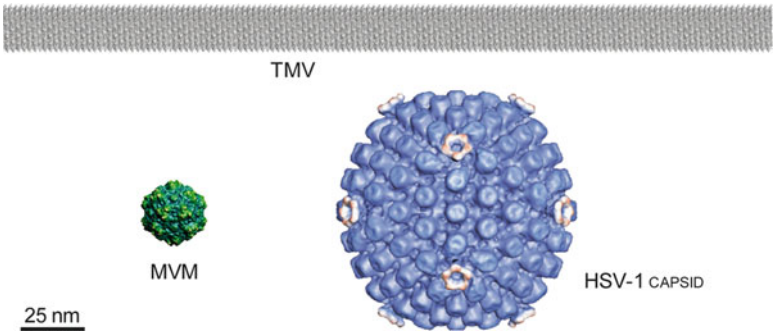


Fig. 1.2 Major types of viral capsid symmetry. *Top image*: helical; the capsid of TMV is shown. *Bottom images*: icosahedral; *left image*, the simple icosahedral capsid of the parvovirus MVM; *right image*, the complex icosahedral capsid of the herpesvirus HSV-1. Images of TMV, MVM and HSV-1 are respectively adapted or reproduced from [Clare DK, Orlova EV (2010) J Struct Biol 171:303–308], [Agbandje-McKenna M, Llamas-Saiz AL, Wang F, Tattersall P, Rossmann MG (1998) Structure 6:1369–1381], and [Grünewald K, Desai P, Winkler DC, Heymann JB, Belnap D, Baumeister W, Steven AC (2003) Science 302:1396–1398], with permission. (Figure kindly provided by M.A. Fuertes)

whose structures have been determined by X-ray crystallography (Chap. 4) are available at the Protein Data Bank (PDB) (http://www.rcsb.org).

Helical Capsids

Helical capsids (Fig. 1.2, top) are extremely simple, theoretically infinite multimers and (in principle) could be made as long as required to encapsidate a nucleic acid genome irrespective of its length; thus, there is no theoretical limitation on the amount of genetic information these capsids could store. However, physical and biological restrictions limit the length of helical capsids, which are much less frequent than icosahedral capsids; regular helical capsids are found in about 10 % of virus families. Chapter 2 provides a description of the architecture of helical capsids; examples of helical viruses can be found in other chapters of this book.

Icosahedral Capsids

In contrast to helical capsids, strictly icosahedral capsids (Fig. 1.2, bottom left) must be made of exactly 60 structurally identical components (CP monomers or oligomers) in order to fulfill intrinsic geometric constraints. This fact could severely limit the size of the nucleic acid genome and, hence, the amount of genetic information that could be enclosed (see Chap. 2). However, evolution has led to different structural solutions to make very large viral capsids with icosahedral symmetry, made of hundreds of chemically identical CP subunits, that can enclose

very large genomes (Fig. 1.2 bottom right). Many large capsids containing many subunits are made of only one type of CP that is capable to adopt different *quasi-equivalent* (similar) conformations (given by the *triangulation number* T); this feature allows the CP molecule to fit into T non-equivalent positions in the icosahedral capsid (see Chap. 2 for a detailed description). This evolutionary strategy minimizes the amount of genetic information required to encode the capsid. Icosahedral capsids are extremely frequent; they occur in about half the virus families. See Chap. 2 for an in-depth description of the architectures of icosahedral viral capsids; Chaps. 2, 10, 11 and others in this book provide many examples of icosahedral capsids and viruses. A database (VIPERdb) contains abundant structural information on viruses with icosahedral capsids whose structure has been determined to high resolution (http://viperdb.scripps.edu).

Virion Architecture

In non-enveloped viruses with a helical or icosahedral capsid, the basic architecture of the capsid determines also the basic architecture of the virion. In enveloped virions, the situation is more complex. Many enveloped virions containing icosahedral capsids and/or other types of compact (nucleo)protein complexes tend to adopt a relatively flexible, frequently spheroidal shape which may differ in size between individual particles.

For example, in herpesviruses (*e.g.*, herpes simplex virus type 1, HSV-1) a large icosahedral capsid (Fig. 1.2, bottom right) containing the viral genome is enclosed in an outer layer of proteins (*tegument*), surrounded in turn by a lipidic envelope; the virion that results is considerably larger than the capsid and spheroidal in shape. In influenza virus (Fig. 1.1, right), several roughly helical nucleocapsid complexes (ribonucleoprotein particles) are directly surrounded by a protein layer (the *matrix*) and the lipid envelope, and this virus is clearly pleomorphic. Many other variations in virion architecture, some of them very complex, do exist. In some cases, two concentric capsids are found; in others, internal lipidic envelopes are present. The structures of some complex viruses are described in detail in Chaps. 2, 11. Other examples of complex virion architectures can be found in several chapters in this book.

1.2.3 Types of Viral Nucleic Acid

The type of genomic nucleic acid used by different viruses is of prime importance to determine the mechanisms of genome replication and expression during the metabolically active phase of the viral cycle. From the perspective of this book the type of nucleic acid, especially whether it is single-stranded (ss) or double-stranded (ds), is extremely relevant for the assembly of many virus capsids (Chaps. 2, 10, 11, 19), the mechanism of nucleic acid packaging (Chap. 12), the organization of the nucleic acid inside the capsid (Chap. 12), virus particle maturation (Chap. 13), and some properties and functions of the virion (Chaps. 2, 10, 11, 12, 13 and 18).

RNA Viruses

Riboviruses (*RNA viruses*) use RNA as genetic material. Some of them use ssRNA (*ssRNA viruses*), others use dsRNA (*dsRNA viruses*). In turn, two types of ssRNA viruses can be distinguished depending on the polarity of their RNA genome strand (*ssRNA*(+) *viruses* and *ssRNA*(−) *viruses*). A genomic single-stranded nucleic acid is considered of positive (+) polarity if its sequence corresponds to that in the viral messenger RNAs (mRNA), and of negative (−) polarity if its sequence corresponds to the complementary of the mRNA sequences. In some viruses, the virion encloses more than one copy of the genome (*e.g.*, retroviruses), or the genome is *segmented*, split into several nucleic acid molecules (*e.g.*, influenza virus). Most of the virus species known are RNA viruses.

DNA Viruses

Deoxyviruses (*DNA viruses*) use DNA as their genetic material. Some of them use ssDNA (*ssDNA viruses*); others use dsDNA (*dsDNA viruses*).

1.2.4 Host Cells and Organisms

Viruses can be grouped also by the kind of cell they can infect (Table 1.1). Many known viruses infect eukaryotic cells (*eukaryotic viruses*). Of them, many infect animals such as vertebrates (including humans) or insects (*animal viruses*); others infect fungi (*fungal viruses* or *mycoviruses*), plants (*plant viruses*) or protista (*protist viruses*). Many other viruses infect prokaryotic cells such as bacteria (*prokaryotic viruses*). Bacterial viruses are usually called *bacteriophages* or, simply, *phages*. Archea are also infected by viruses (*archeal viruses*). Most viruses have evolved to infect one or a few species of organisms (limited *host range*) and one or a few cell types (specific *tropism*). The type of cell a virus can infect is, in part, the consequence of the structural and functional features the virus particle has evolved to enter a particular cell. In general, infection by an animal virus or bacteriophage depends largely on the receptor molecule(s) on the cell surface that the virion can specifically bind (see Sect. 1.4.5 and Chaps. 15, 16, 17). Although some archeal, protist and fungal viruses are providing fascinating cases for structural study (a few of which are mentioned in this book), the vast majority of viruses that have been the subject of structural and molecular virology studies infect animals, plants or bacteria (see Table 1.1). Detailed information on the molecular biology of animal viruses, plant viruses and/or phages can be found also in the books listed at the end of this chapter [1–10].

Table 1.1 Some virus families and species grouped according to type of host and viral genome

Host[a]	Viral genome[b]	Virus order[c]	Virus family[d]	Some virus species[e]
Bacteria	dsDNA		*Corticoviridae*	PM2
		C	*Myoviridae*	P1, P2, P4, SPO1, T4, μ, ϕ92, 8a, 44RR
		C	*Podoviridae*	BPP-1, K1F, P22, Sf6, T7, ε15, ϕ15, ϕ29
		C	*Siphoviridae*	HK97, p2, SF6, SPP1, TP901-1, T5, λ
			Tectiviridae	PRD1, P23-77
	ssDNA		*Microviridae*	ϕX174
			Inoviridae	fd, M13
	dsRNA		*Cystoviridae*	ϕ6, ϕ12
	ssRNA(+)		*Leviviridae*	MS2, Qβ
Archea	dsDNA		*Ampullaviridae*	ABV
			Guttaviridae	SNDV
			Lipothrixviridae	AFV-1
			Rudiviridae	SIRV-2
Protista	dsDNA		*Phycodnaviridae*	PBCV-1
Plant	ssDNA		*Geminiviridae*	BGYMV
	ssRNA(+)		*Bromoviridae*	AMV, BMV, CCMV, CMV
		T	*Flexiviridae* (*Alphaflexiviridae*)	PapMV
			Potyviridae	BYMV
		P	*Secoviridae*	BPMV
			Sobemovirus (*unassigned genus*)	RYMV, SBMV
			Tombusviridae	CPMV, RCNMV, STNV, TBSV, TNV A
		T	*Tymoviridae*	PhMV, TYMV
			Virgaviridae	TMV
	dsDNA-RT		*Caulimoviridae*	CaMV
Fungi	dsRNA		*Partitiviridae*	*Atkinsonella hypoxylon* virus
			Chrysoviridae	*Penicillium chrysogenum* virus
Animal	dsDNA		*Adenoviridae*	(h)Ad
			Asfarviridae	ASFV
			Baculoviridae	AcNPV, cytoplasmic polyhedrosis virus, WNPV
			Iridoviridae	Frog virus 3
		H	*Herpesviridae*	HCMV, HSV-1, VZV
			Mimiviridae	*Acanthamoeba polyphaga* mimivirus
			Papillomaviridae	BPV, HPV 1
			Polyomaviridae	SV40, (human) polyomavirus
			Poxviridae	Vaccinia virus, Variola virus
	ssDNA		*Circoviridae*	PCV2
			Parvoviridae	AAV-2, CPV, FHV, H1-PV, MEV, MVM, PPV

(continued)

Table 1.1 (continued)

Host[a]	Viral genome[b]	Virus order[c]	Virus family[d]	Some virus species[e]
	dsRNA		*Birnaviridae*	IBDV
			Reoviridae	BTV, Rotavirus (A)
	ssRNA(+)		*Astroviridae*	Mamastrovirus 1
			Caliciviridae	NV (Norovirus), RHDV
		N	*Coronaviridae*	SARS virus
			Flaviviridae	Dengue virus, HCV, TBEV, yellow fever virus
			Hepeviridae	HEV
			Nodaviridae	FHV, PaV
		P	*Picornaviridae*	Coxsackievirus, echovirus, FMDV, HAV, HRV, PV
			Tetraviridae (*Alphatetraviridae*)	NωV
			Togaviridae (*genus Alphavirus*)	Chikungunya virus, Semliki Forest virus, Sindbis virus
	ssRNA(−)		*Arenaviridae*	LCMV
			Bunyaviridae	Bunyamwera virus, Hantaan virus, RVFV
		M	*Filoviridae*	Ebola virus, Marburg virus
			Orthomyxoviridae	Influenza virus (A, B, C)
		M	*Paramyxoviridae*	HeV, hMPV, hPIV, measles virus, mumps virus, NDV, NiV, Sendai virus, RSV(Respiratory Syncytial virus)
		M	*Rhabdoviridae*	Rabies virus, VSV
	ssRNA(+)-RT		*Retroviridae*	BLV, EIAV, HIV-1, HIV-2, HTLV-I, MLV, MMTV, MoMuLV, MPMV; RSV(Rous sarcoma virus), SIV
	dsDNA-RT		*Hepadnaviridae*	HBV

[a]Classified according to kingdom. Different virus species of some families may infect organisms from different kingdoms. For example, some species of the *Reoviridae*, *Bunyaviridae*, *Rhabdoviridae* may infect animals and other species may infect plants

[b]*RT* indicates that virus replication involves a reverse transcriptase

[c]The six virus orders are: *C Caudovirales*, *H Herpesvirales*, *M Mononegavirales*, *N Nidovirales*, *P Picornavirales*, *T Tymovirales*. All other virus families listed have not been assigned to any order (see text)

[d]The virus families listed include nearly all of those mentioned in this book and a few others. The current ICTV virus classification includes 94 families, and several independent genera that have not been ascribed to families (see text). The families *Flexiviridae* and *Tetraviridae* have recently been split into several families each; according to the current ICTV classification, the virus species listed for those families belong to the *Alphaflexiviridae* and *Alphatetraviridae* families, respectively

[e]The virus species listed as examples include nearly all of those mentioned in this book and a few additional ones. Virus species abbreviations: *AAV* adeno-associated virus, *ABV Acidianus* bottle-shaped virus, *AcNPV Autographa californica* nuclear polyhedrosis virus, *AFV Acidianus* filamentous virus, *AMV* alfalfa mosaic virus, *ASFV* african swine fever virus, *BGYMV* bean gold yellow

(continued)

Table 1.1 (continued)

mosaic virus, *BLV* bovine leukemia virus, *BMV* brome mosaic virus, *BPMV* bean pod mottle virus, *BPV* bovine papillomavirus, *BTV* bluetongue virus, *BYMV* barley yellow mosaic virus, *CaMV* cauliflower mosaic virus, *CCMV* cowpea chlorotic mosaic virus, *CMV* cucumber mosaic virus, *CPMV* cowpea mosaic virus, *CPV* canine parvovirus, *EIAV* equine infectious anemia virus, *FHV* flock house virus, *FMDV* foot-and-mouth disease virus, *FPV* feline parvovirus, *H1-PV* parvovirus H1, *(h)Ad* (human) adenovirus, *HAV* hepatitis A virus, *HBV* hepatitis B virus, *HCMV* human cytomegalovirus, *HCV* hepatitis C virus, *HeV* Hendra virus, *HEV* hepatitis E virus, *HIV* human immunodeficiency virus, *hMPV* human metapneumovirus, *HPV* human papillomavirus, *hPIV* human parainfluenza virus, *HRV* human rhinovirus, *HSV* human herpesvirus, *HTLV* human T-lymphotropic virus, *IBDV* infectious bursal disease virus, *LCMV* lymphochoriomeningitis virus, *MEV* mink enteritis virus, *MLV* murine leukemia virus, *MMTV* mouse mammary tumor virus, *MPMV* Mason-Pfizer monkey virus, *MVM* minute virus of mice, *NDV* Newcastle disease virus, *NiV* Nipah virus, *NV* Norwalk virus, *NωV Nudaurelia capensis* ω virus, *PapMV* papaya mosaic virus, *PaV* Pariacoto virus, *PBCV Paramecium bursaria Chlorella* virus, *PCV* porcine circovirus, *PhMV Physalis* mottle virus, *PPV* porcine parvovirus, *PV* polio virus, *RCNMV* red clover necrosis mosaic virus, *RHDV* rabbit haemorrhagic disease virus, *RSV* respiratory syncytial virus, *RSV* Rous sarcoma virus, *RVFV* Rift Valley fever virus, *RYMV* rice yellow mottle virus, *SARS* virus, severe acute respiratory syndrome virus, *SBMV* southern bean mosaic virus, *SIRV Sulfolobus islandicus* rod-shaped virus, *SIV* simian immunodeficiency virus, *SNDV Sulfolobus neozelandicus* droplet-shaped virus, *STMV* satellite tobacco mosaic virus, *STNV* satellite tobacco necrosis virus, *SV40* simian virus 40, *TBEV* tick-borne encephalitis virus, *TBSV* tomato bushy stunt virus, *TMV* tobacco mosaic virus, *TNV* tobacco necrosis virus, *TYMV* turnip yellow mosaic virus, *VSV* vesicular stomatitis virus, *VZV* varicella-zoster virus, *WNPV Wiseana* nuclear polyhedrosis virus

Animal Viruses

Animal viruses are extremely diverse in terms of size, shape, presence or absence of envelope, capsid architecture, type of nucleic acid genome, and specific mechanisms employed to complete the different stages of the viral cycle as required for infection, multiplication and propagation. Some animal viruses are among the structurally simplest viruses known. These include the parvoviruses (*e.g.*, the minute virus of mice, MVM (Fig. 1.1, left) or the adeno-associated viruses, AAV; see Chap. 10). Others are structurally much more complex, such as the adenoviruses, herpesviruses (*e.g.*, HSV-1), retroviruses (*e.g.*, the human immunodeficiency virus type 1 (HIV-1) or the giant mimivirus (see Chap. 11)). The study of the structure, function, biology and pathogenicity of many animal viruses has been or is still hampered by a number of difficulties, including their structural and/or functional complexity, technical problems to grow and/or manipulate them and/or safety issues. However, despite all the difficulties, and mainly because of the biomedical or socioeconomic importance of a large number of animal viruses that are pathogenic for humans or livestock, many of these are among the most intensively studied viruses of all. This is reflected also in the many human and animal viruses used as examples and case studies in most chapters of this book. Many particular aspects related to the molecular biology and structure of many different animal viruses can be found also in the books listed at the end of this chapter [1–6, 9–16].

Bacteriophages

Bacteriophages show widely diverse structures and types of nucleic acid genomes; they have helical or icosahedral capsids, and may or may not include a lipid envelope. Phages range from very small and simple non-enveloped icosahedral viruses (*e.g.*, ϕX174) and long but simple helical viruses (filamentous phages), to large and complex viruses (*e.g.*, tailed DNA phages such as ϕ29; Fig. 1.1, center). Since the origins of Molecular Biology over half a century ago, and continuing through several decades, some bacteriophages were found to present important advantages as model systems for molecular and genetic studies compared to most animal and/or plant viruses. Their advantages include the facility to grow phages and their bacterial hosts in large amounts; the relative structural and functional simplicity and ease of handling of bacterial cells compared to eukaryotic cells; and the possibility to readily obtain certain types of mutant viruses to investigate virus structure and function. These and other reasons led to the intensive use of phages as paradigms for many molecular biology studies on nucleic acid replication, gene expression and their regulation. Several bacteriophages have also provided and continue to provide model systems for studying molecular recognition and self-assembly during the morphogenesis of biomolecular complexes. As a consequence, the structure and function of some bacteriophages, and most stages in their life cycles, are known in great detail (see Chaps. 11, 17). Many particular aspects of the molecular biology of different bacteriophages can be found in several books listed at the end of this chapter. For a book on bacteriophage molecular biology see [7].

Plant Viruses

Most of the very abundant plant viruses are non-enveloped ssRNA(+) viruses with a slender helical capsid or a relatively small icosahedral capsid. The structure and/or function of a few plant viruses, such as the tobacco mosaic virus (TMV; Fig. 1.2 top) have been intensively studied for many decades, in many cases because of some advantages of those viruses as model systems; for example, the facility to grow plant viruses in very large quantities by simply infecting host plants. In addition, studies on many plant viruses have been greatly stimulated because of the economically important diseases they cause in crop plants. Recently, some of the advantages of phages and plant viruses referred to above have led to their preferential use as platforms for many bio/nanotechnological developments (see Chap. 22). However, generally speaking plant viruses have been the subject of fewer studies than animal viruses or bacteriophages, and many stages of the life cycles of the former remain less well known than those of the latter. For a book on plant virus molecular biology and structure see [8].

1.2.5 Classification of Viruses

Comparisons of the sequences of viral genes and genomes have led to the establishment of phylogenetic relationships between many viruses. In addition, comparison of the tertiary structure of viral proteins, especially CPs, has allowed the tentative proposal of distant evolutionary relationships among different viruses, or at least between some of their genes (see a brief description in Chap. 7). It must be noted here that genetic recombination and horizontal gene transfer between even very different, unrelated viruses are frequent. Thus, viruses of widely different origins could share some evolutionarily closely related genes, and of course the proteins these genes encode. It is not yet possible to solidly establish a general phylogenetic-based classification of viruses. In 1973 the International Committee on Taxonomy of Viruses (ICTV) was established, and a general database on viruses was created later (http://www.ncbi.nlm.nih.gov/ICTVdb).

Known viruses have been classified into seven major groups based on the type of nucleic acid genome (*Baltimore classification*). These groups are: I: dsDNA viruses; II: ssDNA viruses; III: dsRNA viruses; IV: ssRNA(+) viruses; V: ssRNA (−) viruses; VI: ssRNA(+) virus whose replication involves the action of a reverse transcriptase (RT) that synthesize DNA from a RNA template; VII: dsDNA viruses whose replication involves the action of an RT (see Table 1.1).

In addition to their classification according to type of nucleic acid genome, viruses have been classified by ICTV in a number of taxonomic groups (*taxons*): viral *order*, *family*, *subfamily*, *genus*, and *species*. The most useful taxon in virus classification is the *family* (see Table 1.1). Viruses in a same family probably share a not too distant evolutionary relationship, as established mainly by comparative sequence analysis. There are currently 94 recognized virus families; only 22 of these families have been grouped in 6 orders (*Caudovirales*, *Herpesvirales*, *Mononegavirales*, *Nidovirales*, *Picornavirales*, *Tymovirales*; see Table 1.1); the remaining families have not been assigned to any order yet. Also, several virus genera have not been assigned to any family yet.

Virus family latin names (italicized) include the suffix –*viridae*. Very frequently, the family english name (non-italicized), which include the suffix –virus (plural - viruses) is used instead of the latin name. However, this may occasionally cause confusion on whether one is referring to a viral family or genus, unless this point is specified. Virus species are usually referred to by their english names, and most have been given standard abbreviations. Most bacteriophages are named according to a code of latin letters, greek letters and/or numbers.

In this book, viruses will be generally identified by type of nucleic acid, by family (latin or english name) and/or by species. For example, in different chapters repeated mention is made to the human immunodeficiency virus type 1 (HIV-1). HIV-1 is a virus species of the *Retroviridae* (retrovirus family), which belongs to a group of ssRNA(+) viruses whose replication involves a RT (group VI). A list of

nearly all virus families and species mentioned in this book, along with a few other important families and species, is included in Table 1.1. This table is mainly intended to help the reader navigate among the multitude of virus names that will inevitably appear along the different chapters of this book (or of any other virology book). In addition, an index of names of nearly all virus species and families mentioned in this book can be found at the end of the book.

Viruses in each family do share many features and present many structural and functional similarities, apart from the type of nucleic acid genome. Of the utmost relevance for the subject of this book is that viruses from a same family generally show a clear similarity regarding virion and capsid structures. However, it must emphatically be noted also that even viruses of a same family may sometimes *drastically* differ in some or many structural features, properties, functions and underlying mechanisms, cells and organisms they can infect, pathogenic effects, etc. Substantial phenotypic differences may occur even between some individual viruses from a same clone, and some of these differences can be due to a single amino acid substitution in some viral protein. The converse is also true: some viruses from different families may share many structural and functional similarities, probably as an evolutionary consequence of similar "lifestyles" (either by conservation of basic structural and/or functional features, or by convergent evolution; see, for example, Chaps. 2, 7 regarding structural similarities in CPs and Chaps. 2, 19 for fundamental physics-based similarities in architecture and some properties of icosahedral virus capsids). Figure 1.3 shows very simple general schemes of virion structures for members of some important families of animal viruses, many of which are mentioned in this book.

1.3 Techniques Used to Study the Structure and Physics of Viruses

Today, the structural virologist has at his/her disposal a vast array of sophisticated and powerful techniques to study virus structure, dynamics (including conformational rearrangements, assembly and disassembly) and physicochemical properties. Most of the major structural and many biophysical experimental techniques in current use to study isolated viruses are covered in Part II (Chaps. 3, 4, 5, 6, 7, 8 and 9). Several experimental techniques used to study viral structural components and complete virus particles inside the cell are described in Chap. 14. Several physics-based theoretical approaches to study the structure, dynamics and properties of viruses are covered in Chap. 19.

There are no easy recipes to decide which one(s) of those techniques a researcher should use for solving a specific problem in structural virology. The decision will depend, of course, on many scientific, technical and practical circumstances and considerations. Some options may be easily discarded, but several other options

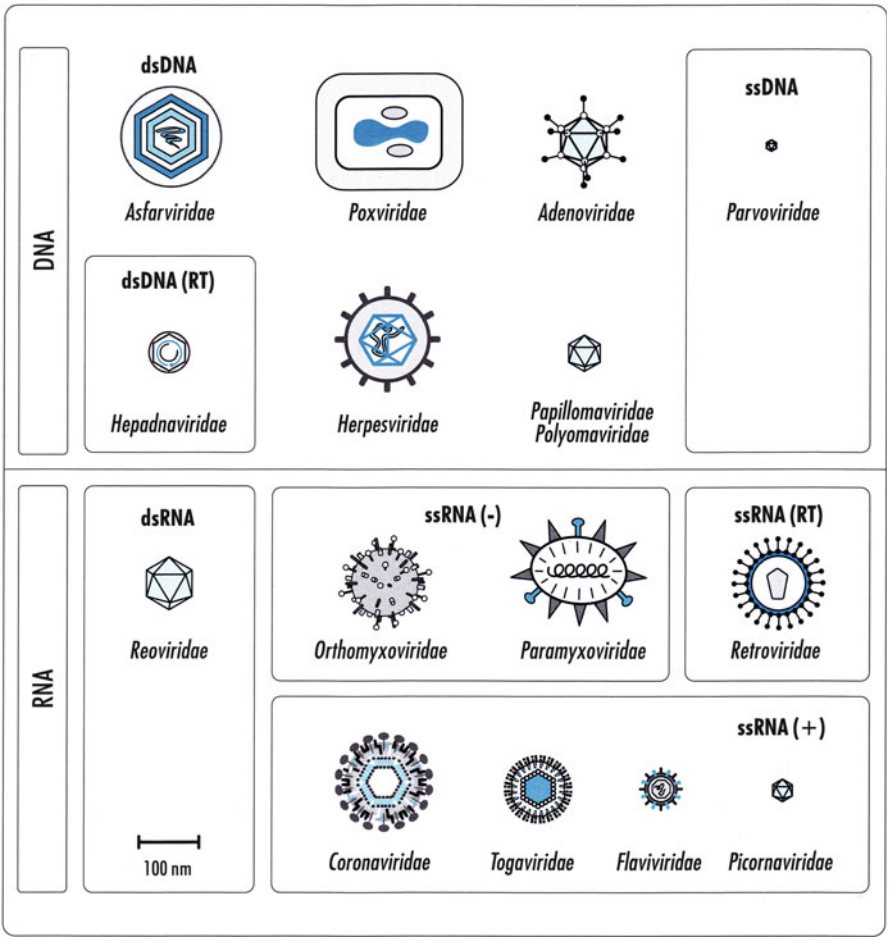


Fig. 1.3 General morphologies of virions belonging to some important animal virus families. The families are grouped according to viral DNA type (Figure reproduced from [Carrasco L, Almendral JM (ed) (2006) Virus Patógenos. Ed. Hélice and Fundación BBVA, Madrid], with permission from Ed. Hélice and Fundación BBVA)

may, in principle, be adequate for any intended goal, model virus and defined situation. Asking oneself a number of questions when considering a structural virology project should greatly help in defining the technical choices. A number of relevant considerations (including a few obvious ones) are listed next:

Some scientific considerations:

(i) What are the researcher's specific scientific goals?
(ii) What information on the chosen virus is already available?

(iii) Is the researcher interested in determining the structure of the complete virion, or only of some structural component(s)? Are the virus particles asymmetrical, or pleomorphic? Are they enveloped?
(iv) Is the researcher interested in studying some dynamic process in which the viral particle could be involved? Dynamic studies may be focused on fluctuations of the viral structure, or externally induced conformational rearrangements, including the structural characterization of intermediates of virus particle assembly or disassembly.
(v) Does the researcher contemplate the study of some aspect of virus structure or dynamics inside a host cell?
(vi) Is the researcher interested in characterizing some physicochemical property of the virus particle? Which one(s)? What for?
(vii) Is the researcher interested in determining the structure or properties of a virus particle or component bound to a ligand? Which one?
(viii) Is the researcher interested in trying to relate the structural information to be acquired on the viral particle or component or virus-ligand complex with some viral function? Which one(s)? Does the researcher aim at identifying the molecular mechanism involved?
(ix) Is the researcher interested in trying to relate the structural information to be acquired with the molecular biology of the virus? With pathogenic effects? With viral evolution? Any other aspect?
(x) Does the researcher intend to develop any particular application based on the structural information acquired? Which application?

Some technical considerations:

(xi) Can the virus in question be produced by infection of cultured cells? or can the viral CPs be produced in some expression system, and will they form viral capsids or virus-like particles (VLPs)? can isolated viral components of interest be obtained in a stable form?
(xii) Can the virions, capsids or VLPs be purified? Which level of purity must be reached? What amounts of sufficiently pure viral particles or components can be reasonably obtained?
(xiii) Do the viral particles or components aggregate at high concentrations? Can they be crystallized?
(xiv) Are the viral particles or components stable enough in the conditions required for structure determination by the technique(s) being considered?
(xv) Should the viral particles or components be chemically or genetically modified? Is there an infectious DNA clone of the virus genome available to allow site-directed genetic modifications?

A few practical considerations:

(xvi) Which structural virology expert(s), if any, could be contacted for possible collaboration? What structural facilities are available to the researcher and/or his collaborators?

(xvii) What about safety issues? Will containment facilities be available, if required? Permissions needed?

(xviii) How much time and what resources will be available for the project? What partial goals should be achieved along the project? In which order should they be tackled? By whom? How long will each goal possibly take? What are the alternatives if something fails or takes (much) more time than expected? These latter considerations are indeed common to all scientific projects, but the very complex nature of many structural techniques applied to study a large molecular assembly, and the time, effort and uncertainties involved, make such considerations particularly relevant in this case (and difficult to decide upon).

Given the very large number of considerations and uncertainties, it would be illusory to try and provide some general guidelines for choosing the "right" technique(s) for analysing virus structure, dynamics or physical properties. However, a brief comparison on the general applicability, strengths and limitations of each of the different structural techniques for different generic studies in structural and physical virology is possible and is included next. Relevant examples on the applications of the different structural and physical techniques to specific problems in virology are provided in most chapters of this book.

1.3.1 Techniques for Studying the Structure of Virus Particles or Their Components

Several powerful techniques are being used to determine the high-resolution structure of virions or their structural components. The detailed structures of many different virus particles have provided both critical answers and relevant new questions in molecular virology, and have led to a spectacular advance in our knowledge of viruses and biological processes, and our options to combat viral disease.

Transmission Electron Microscopy Using Negative Staining

Conventional *transmission electron microscopy* (EM) (Chap. 3) remains an all-important imaging technique, especially for initial structural characterization of isolated virus particles, and for many in-cell studies of virions or their components (Chap. 14). It generally requires very small amounts of virus particles that do not need to be very pure, and is extremely fast. The particle general morphology and architecture can be determined.

However, the sample must be fixed and covered (stained) with a heavy metal layer (opaque to electrons) to increase contrast. This treatment may introduce artifacts, including deformation of flexible particles, alterations of their fine

structure and blurring of their images. The resolution achieved on stained viral particles is, in general, relatively limited. In addition, only the surface topography of the particle can be imaged (see Chap. 3).

Cryo-Electron Microscopy

Cryo-electron microscopy (cryo-EM) (Chap. 3) is a variant of EM, and one of the most important techniques for imaging isolated virus particles and virus-ligand complexes. The sample is vitrified and kept at extremely low temperature; the virus particles remain hydrated, and no staining is generally required. The possibility of artifacts is greatly reduced compared to conventional EM. Even more important, cryo-EM is invariably used together with three-dimensional reconstruction techniques that average the images of thousands of virus particles visualized in different orientations, and result in a volumetric (three-dimensional) structural model of the imaged particles. Because the images of so many individual particles are averaged, different artifacts affecting different individual particles make very little contribution to the averaged model, and the signal-to-noise ratio is greatly increased. Also, as no staining agent needs to be used, the structural model obtained reveals also the anatomy (internal structure) of the viral particle. The typical resolution achieved on virus particles was, until relatively recently, about 1 nm or less, but in the last years cryo-EM models of some virus particles with a resolution around 0.5 nm, or even higher, have been obtained, thus approaching atomic resolution. In these high-resolution models, the tracing of the polypeptide chains in the viral capsids and other fine structural details, including the orientation of some amino acid side chains, can be observed (see Chap. 3). Cryo-EM has been repeatedly used also to obtain the structures of virus particle-ligand complexes which have not been generally crystallized and whose structure, thus, could not be directly determined by X-ray crystallography.

However, in cryo-EM thousands or tens of thousands of good-quality images must be taken, processed and analyzed; all things considered, a cryo-EM project may sometimes take months to be completed. Reaching a very high resolution is not guaranteed, especially with complete virus particles and other large biomolecular complexes. Because the process involves image averaging, the technique is best applied to symmetric viruses, and is not suitable for pleomorphic viruses (see Chaps. 3, 7).

Cryo-Electron Tomography

Cryo-electron tomography (Cryo-ET) (Chap. 3) is a variant of cryo-EM in which an individual microscopic object can be imaged using various angles of view. As a result, a three-dimensional reconstruction of the structure of a single virus particle (even inside a cell, see Chap. 14), and not a reconstruction based on many different particles, can be achieved. Because of this feature, cryo-ET is able to image pleomorphic virions such as influenza virus (Fig. 1.1, right), retroviruses and

many others, for which cryo-EM together with three-dimensional reconstruction cannot be used. A major disadvantage of cryo-ET is the much lower resolution achieved (say, about 3 nm) when compared to cryo-EM (see Chap. 3).

X-ray Crystallography

X-ray crystallography (Chap. 4) is a crucially important technique for the determination of the high-resolution structure of both virus particles and their isolated molecular components. Atomic or near-atomic resolution is generally achieved and there are, in principle, no limitations in the size or shape of the particle whose atomic structure is to be determined. The structures of very large and structurally complex virions and capsids, and of very large and asymmetric biomolecular complexes such as the ribosome have been already determined by X-ray crystallography. Molecular crystals contain a large proportion of water, the virus particles or proteins in them are fully hydrated, and the structures obtained are physiologically relevant.

A major disadvantage is that adequate molecular crystals of the virus particles or their components must be obtained, which is not always easy, or even achieved at all. For example filamentous, enveloped, or pleomorphic viruses will not crystallize; adequate crystals of flexible multimeric complexes between virus particles and large protein ligands such as cellular receptors or antibody fragments have very rarely been obtained. Also, relatively large amounts of very pure virus particles or proteins are most often required. Crystals must be well-ordered, of sufficient size, and stable enough under an intense X-ray beam to allow the collection of adequate data. The question of how to determine some necessary data (the *phase problem*, see Chap. 4) must be considered. In general, determination of the crystal structure of a virus particle, or even of a viral subassembly or individual protein is not an easy task and the project may take many months, or even several years.

Nuclear Magnetic Resonance (NMR) Spectroscopy

Solution NMR spectroscopy (Chap. 5) complements X-ray crystallography for the determination of the atomic or close to atomic resolution structure of molecular components of virus particles. The molecules remain in solution, generally in close to physiological conditions, and the uncertainties of the crystallization process are avoided. The major disadvantage of NMR for structural determination is a rather severe limitation in the size of the biomolecule whose structure can be solved. At present, molecules for structural determination by NMR must be no larger than about 35 kDa. However, by using state-of-the-art instrumentation and sophisticated technical approaches the structures of some larger proteins (50 kDa and more) have been solved. Unfortunately, direct structural determination of complete viral particles, most viral subassemblies and many multidomain or oligomeric proteins are still way beyond the reach of solution NMR spectroscopy. Large amounts of

purified sample and, usually, isotopic labelling of the sample (see Chap. 5) are required too, and the biomolecules must not aggregate during the slow acquisition of data at moderate temperature. *Solid-state NMR spectroscopy* has been applied for structural studies of a few viral particles (see Chap. 5).

Atomic Force Microscopy

In *atomic force microscopy* (AFM), small objects are not probed by electrons or photons as in EM, X-ray crystallography or NMR spectroscopy. Instead, the objects are probed by "touching" them with a very fine stylus (tip) that is used to trace the topography of the object (Chap. 8). AFM has been already applied to image a number of virus particles (see Chap. 8). Its greatest advantage as a static virus imaging technique is that the structure of single particles can be determined in liquid, in close to physiological conditions. Most importantly, AFM can also be used as a *dynamic* imaging technique, where structural changes can be followed in real time (see below). A clear disadvantage is that, like conventional EM but unlike cryo-EM, cryo-ET, X-ray crystallography and NMR spectroscopy, only the surface topography of the object (not its internal structure), can be determined. Because of basic geometric considerations regarding the tip-sample interaction, care has to be exercised in interpreting the images obtained (see Chap. 8). Also, flexible particles could be deformed by adsorption to the solid substrate. The resolution on virus particles is currently limited to about 1 nm at best. In special conditions (ultra-high vaccuum) with special instrumentation, atomic resolution has already been achieved using AFM on small organic molecules. See Chap. 8 for a detailed description of this *single-molecule technique* and its application to imaging viruses.

Combined Structural Approaches

Some limitations of X-ray crystallography, NMR spectroscopy and cryo-EM have been overcome by combining these techniques (see Chap. 7). For example, the atomic structures of a virus particle and its receptor molecule can be independently solved by X-ray crystallography, and the lower resolution structure of the virus particle-receptor complex can be determined by cryo-EM; by combining the structural models obtained, a high-resolution pseudo-atomic model of the complex can be obtained. In another example, if a complete viral capsid cannot be crystallized, the atomic structure of the isolated CP subunit can be solved by NMR spectroscopy or X-ray crystallography, and the lower-resolution structure of the complete capsid can be determined by cryo-EM; again, by combining the structural models obtained, a detailed pseudo-atomic model of the complete capsid can be constructed. The power of combining these and other structural, biophysical and biochemical techniques for structural virology studies is fully illustrated in Chap. 7, and can be also appreciated in examples described in other chapters of this book.

1.3.2 *Biophysical Techniques for Studying the Dynamics and Physical Properties of Virus Particles*

The high-resolution "photographs" obtained by applying the techniques just mentioned do provide invaluable structural information; however, there is a clear need to complement those static images with "movies" on the dynamics of virus particles using the highest possible spatial and temporal resolution. This is a difficult task, and the available techniques that have been applied so far to viral particles have generally provided relatively limited views of their dynamics. However the results already obtained have certainly established that virus particles are highly dynamic complexes, and provided abundant evidence on the critical importance of the local and global dynamics of virus particles and/or some of their components during the infection cycle (see Chaps. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19 for many specific examples).

Two types of dynamism in virus particles can be distinguished: equilibrium dynamics (*e.g.*, *breathing*), in which the structure of the virus particle in any particular state fluctuates around a minimum energy conformation; and reversible or (frequently) irreversible conformational transitions between two states of the viral particle, induced by external physical and/or chemical external factors *in vivo* and/or *in vitro* (see Sect. 1.4.4).

In the case of conformational rearrangements, the initial and final states and any reaction intermediates that happen to be stable enough may, in principle, be structurally characterized using the techniques already mentioned in Sect. 1.3.1 and described in detail in Chaps. 3, 4, 5, 7. In addition, other biophysical or biochemical techniques can be used to detect changes in secondary, tertiary and/or quaternary structure in the viral capsid. These techniques include *circular dichroism spectroscopy* and *fluorescence spectroscopy*, and these have been used also to probe in solution some particular aspects of the structure and properties of virus particles, including chemical and conformational stability (see Chap. 6). In some cases, unstable intermediates of dynamic processes involving virus particles have been stabilized by kinetic trapping using biochemical or genetic methods, and characterized using some of the structural techniques referred to above. A few transient intermediates of assembly/disassembly have been identified by using *mass spectrometry* (MS) techniques (Chap. 6). Small-angle X-ray scattering (SAXS) (Chap. 7) can be used to determine the shape of virus particles and contributes to structural determination in combination with other techniques; in addition, SAXS has started to be used to study virus dynamics including assembly and disassembly (see Chap. 7). *Differential scanning calorimetry* can also be used to follow conformational transitions in virus particles (see Chap. 7).

Single-molecule physical techniques such as AFM (Chap. 8) and *optical tweezers* (Chap. 9) have recently begun to be applied to study dynamic processes in viral particles. AFM offers the possibility not only to structurally characterize initial and final states and stable intermediates in dynamic processes, but to follow in real-time conformational rearrangements or the disassembly of viral particles, and to probe

some of their physical properties (see Chaps. 8, 18). Optical tweezers has been used to follow and quantify in real-time the dynamics of DNA packaging in bacteriophage capsids (see Chap. 9).

Equilibrium dynamics of a number of viral proteins has been probed by *NMR spectroscopy* (see Chap. 5). Global or local equilibrium dynamics (breathing) of complete virus particles has been probed by different biophysical or biochemical techniques including *NMR spectroscopy* (Chap. 5), *limited proteolysis/MS* (Chap. 6) and *hydrogen exchange/MS* (Chap. 6).

Physics-based models and computational simulations, including *molecular dynamics* (MD) and other approaches are being used to predict some aspects of virus particle dynamics. However, at present even the smallest virus particles are too complex for long enough all-atom MD simulations; thus, highly simplified (coarse-grained) structural models of simple viral capsids are used instead. Recently, a few short all-atom MD simulations of some aspect of the dynamics of very simple viral capsids have been carried out. These and other physics-based theoretical approaches to study virus particle dynamics, including conformational changes, assembly and disassembly are described in Chap. 19.

1.3.3 Techniques for Studying Structure-Function Relationships in Viral Particles

Once the molecular structure and/or conformational dynamics of a virus particle or structural component are known, a vast array of biophysical, biochemical, genetic, cell biology and other techniques can be used to investigate the relationship between viral structure and viral properties or function. An account of those techniques and how they can be used to study virus structure, properties and function are outside the scope of this book. However, some of these techniques and their uses in structural virology are briefly mentioned in Chaps. 7, 14. Examples of the results obtained in structure-function analyses of virus particles or their components can be found also in other chapters of this book (see especially Chaps. 10, 11, 12, 13, 14, 15, 16, 17 and 18).

1.4 The Roles of Virus Particles and Their Components Along the Virus Life Cycle

1.4.1 General Overview of the Virus Cycle

Viruses have evolved different alternatives for their multiplication and propagation. However, these alternatives share a number of general features and involve a number of broadly similar processes, albeit with multiple variations.

Several types of virus infection are known; which one may occur will depend on the virus species, host cell and conditions in which the cells are infected. In the case of animal viruses, many infections lead directly to the production of progeny virions in large numbers and the death of the host cell (*lytic infection*). In contrast, some infections involve the constant production of virus particles in small numbers, without killing the infected cells which are able to divide, and both the infected cells and the virus will persist in time (*persistent infection*). Other infections lead to the indefinite perpetuation in the host cell of the viral genome, which is replicated along with the cellular genome, and propagated by passing to the progeny cells (*latent infections*). In these infections, the viral genome may be propagated by becoming physically integrated in the cellular genome, as in the case of HIV-1, or by being maintained as an independent (*episomal*) nucleic acid molecule, as in the case of HSV-1. However, even latent infections can eventually lead to the production and release of virus particles that will be able to infect other host cells, closing the virus life cycle. Some bacteriophages can also produce both lytic and latent infections, which are viewed as evolved alternative strategies for virus survival; either alternative will be followed depending on the conditions the infected cell and the virus may encounter.

A generic, simplified virus life cycle is illustrated in Fig. 1.4. It must be emphasized again that many variations are possible depending on the virus species, host cell and environmental conditions. However, this simplified virus cycle may be useful in this book for putting into perspective the different processes in which virus particles and their molecular components participate during the infection process. The cyclic existence of a virus physical entity and its descent is traditionally contemplated in textbooks from the perspective of an infected host cell. Thus, a viral cycle is generally described by starting from a pre-existing, already formed virus particle which infects a cell; and ending with the progeny virus particles that are eventually produced in, and released from the infected cell (either directly in a lytic infection, or after a number of cellular generations have gone through a latent infection). From this classic perspective, the major generic steps in the life cycle of a virus (especially adapted for animal viruses, but mostly valid also for bacteriophages) can be listed in the following order, from step (1) to step (12) (see Fig. 1.4, where steps are identified using the same numbering):

Cell recognition by the virion and entry

(1) Recognition of receptor(s) molecule(s) (and eventually co-receptors) on the host cell surface.
(2) Internalization (entry) of the virus particle or of some of their components (invariably including the viral genome) into the host cell.
(3) Trafficking of the viral particle or some viral structural component (invariably including the viral genome) in the host cell.

Viral genome uncoating

(4) Release of the viral nucleic acid in some intracellular location where its replication will take place. Unless only the naked nucleic acid genome enters the cell (as in the case of many bacteriophages), release of the nucleic acid from

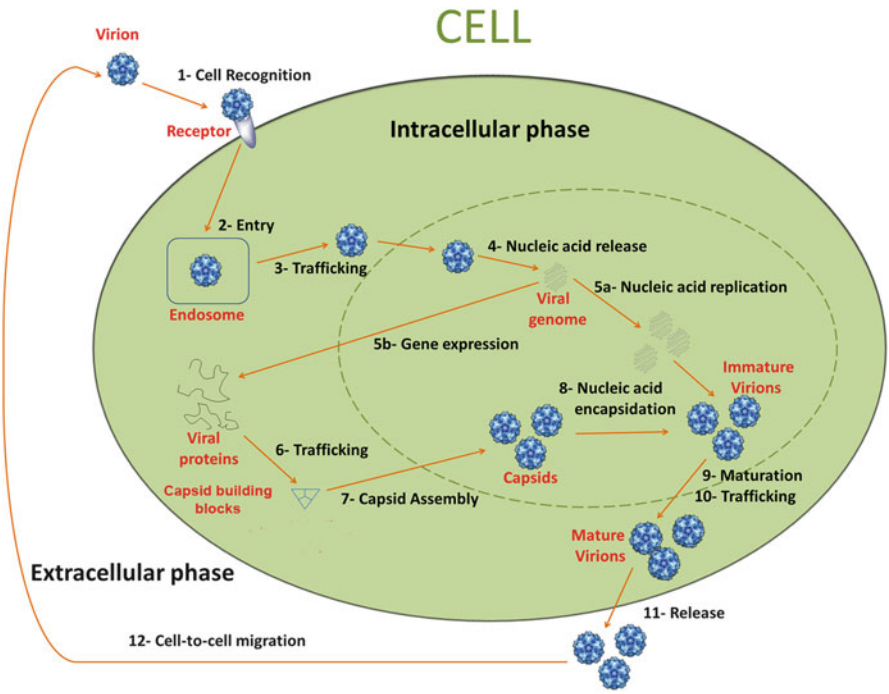


Fig. 1.4 A generic viral cycle. The following stages of the cycle are schematized: *1* Recognition of host cell receptor(s) on the host cell surface. *2* Entry of the virus particle (or of some components) into the host cell. *3* Trafficking of the viral particle (or some components) in the host cell. *4* Release of the nucleic acid. *5a* Replication of the viral genome. *5b* Expression of the viral genome. *6* Trafficking of the synthesized virus components to sites where the morphogenesis of virus particles will take place. *7* Assembly of the virus capsid from the previously synthesized CP subunits. *8* Packaging of the nucleic acid genome inside an assembled capsid (capsid assembly and nucleic acid packaging may occur concomitantly). *9* Maturation of the virus particle. Depending on the animal virus species, stages 4, 5a, 5b (in part), 6, 7, 8 and 9 may occur in either the cytoplasm or the nucleus of the host cell; this is indicated in the figure by encircling those steps by a discontinuous line, which represents the nuclear membrane only in those cases where the processes indicated occur in the cell nucleus. *10* Intracellular trafficking of the virus particles and/or mature virions, which may occur between, during and/or after steps 7, 8, and/or 9. *11* Release of virions from the cell. *12* Navigation of virus particles in the extracellular environment. See text for a more complete description and some important variations (Figure kindly provided by M.A. Fuertes)

the viral particle carrier (viral genome uncoating) must occur there. From this moment until the progeny virus particles are assembled, the virus fully enters the *eclipse phase*. The parental virus loses completely its identity as a molecular complex, and only its genome persist (and, eventually, also some of its viral proteins if required during the intracellular phase of the viral cycle).

Viral gene expression and genome replication

(5) Expression and replication of the viral genome. Unless a latent infection is established, *e.g.*, by integration of the viral genome into the cellular genome, the cellular metabolism is subverted to the virus advantage. Viral genes are transcribed into mRNAs in spatially and temporally controlled processes, and viral proteins required during the different processes of viral metabolism, and structural viral proteins required to form the progeny virus particles are synthesized. In addition, the viral nucleic acid genome is replicated by viral and/or cellular polymerases, following one of a vast array of mechanisms that will depend in part on the type of viral nucleic acid genome and virus species.
(6) Trafficking of the synthesized virus components to intracellular sites where the morphogenesis of virus particles will take place.

Virion morphogenesis

(7) Assembly of the virus capsid from the previously synthesized CP subunits.
(8) Packaging of the nucleic acid genome inside an assembled capsid. For many viruses, capsid assembly and nucleic acid packaging occur concomitantly in a single condensation process.
(9) Maturation of the virus particle to become an infectious virion. For some viruses, maturation occurs in response to nucleic acid packaging, and these processes cannot be clearly separated.

Virion release

(10) Intracellular trafficking of the progeny virus particles and/or mature infectious virions, which may occur between, during and/or after steps 7, 8, and/or 9.
(11) Release of virions from the cell. For some viruses, maturation may occur during, or even after release.

Extracellular existence of the virion

(12) Navigation of virus particles in the extracellular environment, within and between organisms. During this apparently passive phase, the virion is in fact confronted with, and must fight against, multiple external factors that can lead to its disruption or inactivation. These factors include many physical and chemical aggressions in the environment and multiple antiviral biochemical defenses mounted in the host organism: for example, virus neutralization by antibodies and/or other components of the immune system of an animal host.

This book is focused on virus particles, and a major part of it (Part III, Chaps. 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19) deals with each of the major general stages in the virus cycle *in which virus particles or their molecular components participate* (*i.e.*, steps 1–4 and 7–12 above). Thus, we have chosen the "beginning" of the (in fact) endless virus cycle (Fig. 1.4) from the perspective of a virus particle, and not of an infected cell. We start our description after the viral components are synthesized in a host cell and when the viral particle begins its existence by self-assembly from some of those components (Chaps. 10, 11, 12, 13 and 14); and we proceed until the same viral particle ceases to exist by uncoating its nucleic acid

genome in another host cell (Chaps. 15, 16, 17). A logical advantage of this approach is that during morphogenesis of a virion critical features are created that will be needed later by that same virion, for example when it enters the cell or when its genome is uncoated. Examples of such features include metastability of many viruses or their components (Chaps. 10, 11, 12, 13, 15 and 16), or built-in pressure in tailed bacteriophages (Chaps. 9, 12, 17, 19).

1.4.2 Viral Metabolism: Viral Genome Expression and Replication

After a virion infects a host cell, it begins to lose its physical integrity, finally disappearing as such when its nucleic acid is released inside the cell. From this point until the progeny viruses are formed (if they are formed at all), the virus as a discrete physical entity ceases to exist. What we find instead is an infected cell that contains an additional set of (viral) genetic instructions. The viral genome and, in some viruses, a few viral proteins that have entered the host cell along with the genetic material may not only start a number of new metabolic events, but may also alter the cellular metabolic processes and the structural organization of the cell. Eventually, many copies of the viral genome and of the viral proteins required to form a new virus particle are made in the infected cell, and a number of progeny virions are assembled from them.

The diverse viral metabolic processes, such as viral genome replication, recombination, integration, transcription and translation and their regulation, and some of the virus-induced alterations of cellular components and reactions, are generally much better known than most stages in which viral particles participate, such as virus morphogenesis or disassembly and uncoating. Viral metabolism, including some structural and many functional aspects, constitutes the major part of most molecular virology textbooks and is out of the scope of the present textbook, which focuses on the structural biology and physics of virus particles instead. Thus, only a few very general basic facts on viral metabolism will be sketched here. For detailed accounts on viral metabolism the reader is referred at the end of this chapter to a few excellent molecular virology textbooks [1–8], and to the specialized reviews and references contained in them.

Partly because of the different types of nucleic acid molecules that can act as viral genomes, different viruses follow a bewildering diversity of pathways and mechanisms for genome expression and replication. Moreover, genetic information in viral genomes is stored in a remarkably compact ("streamlined") configuration, probably because of size limitations in the viral capsid and other constraints. This fact also contributes to the notorious complexity observed in the regulation of viral gene expression in many viruses.

A central requisite for viral gene expression (as for cellular gene expression) is the synthesis of mRNA molecules (*genome transcription*). The dsDNA genomes of

herpesviruses, adenoviruses, or poliomaviruses (to cite some examples) are replicated by the action of a DNA polymerase in the cell nucleus, and their genes are transcribed into mRNAs by a RNA polymerase much like the genes in cellular chromosomes. In hepadnaviruses, replication of the dsDNA genome involves a RNA intermediate and its reverse-transcription by a RT. The ssDNA genomes of parvoviruses are copied by a DNA polymerase in the nucleus to yield dsDNA molecules, which are transcribed into mRNAs, and also used as intermediates to yield multiple progeny ssDNA genomes that are encapsidated during virus morphogenesis in the nucleus.

ssRNA viruses are extremely abundant and also diverse. In order to be transcribed into mRNAs, viral ssRNA(+) genomes, such as those of picornaviruses or coronaviruses, are replicated in the cytoplasm using a viral RNA replicase to yield complementary ssRNA(−) molecules, which are in turn replicated/transcribed to yield complementary ssRNA(+) molecules that may act as viral genomes and/or mRNAs. Viral ssRNA(−) genomes, such as those of orthomyxoviruses (influenza virus) or paramyxoviruses, are transcribed in the cytoplasm to yield complementary ssRNA(+) molecules that may act as mRNAs and/or are in turn replicated to yield ssRNA(−) molecules that act as viral genomes. In retroviruses, the ssRNA(+) genome is reverse-transcribed by a viral RT into a complementary ssDNA(−) molecule, which is then used as template to form a viral dsDNA molecule which may be integrated in the cellular genome and/or transcribed to yield viral mRNAs. The dsRNA genomes of reoviruses are directly transcribed into mRNAs and replicated in the cytoplasm.

The transcription of viral genes into viral mRNAs inside the infected cell is a carefully regulated process, both in space and time. In many but not all viruses, during a lytic infection some viral genes are preferentially transcribed early in the infection process, and the corresponding mRNAs are translated in the cell ribosomes to yield *early viral proteins*, which will be generally needed to subvert the cell metabolism or control later stages of viral gene expression. At some point, replication of the viral genome starts and genomic nucleic acid molecules accumulate in some location of the infected cell. In later stages, other viral mRNAs are translated to yield *late viral proteins*. These are generally synthesized in much larger amounts than early viral proteins, and usually include, among others, the *structural viral proteins*, those that will form a part of the newly assembled virus particles during virion morphogenesis, the next stage of the viral cycle.

1.4.3 Virion Morphogenesis

Virion morphogenesis is a very complex and tightly regulated process and, except for a few viruses, is still poorly understood in many aspects. Several different stages that sometimes overlap can be distinguished: capsid assembly (Chaps. 10, 11), nucleic acid packaging (Chap. 12) and virus particle maturation (Chap. 13).

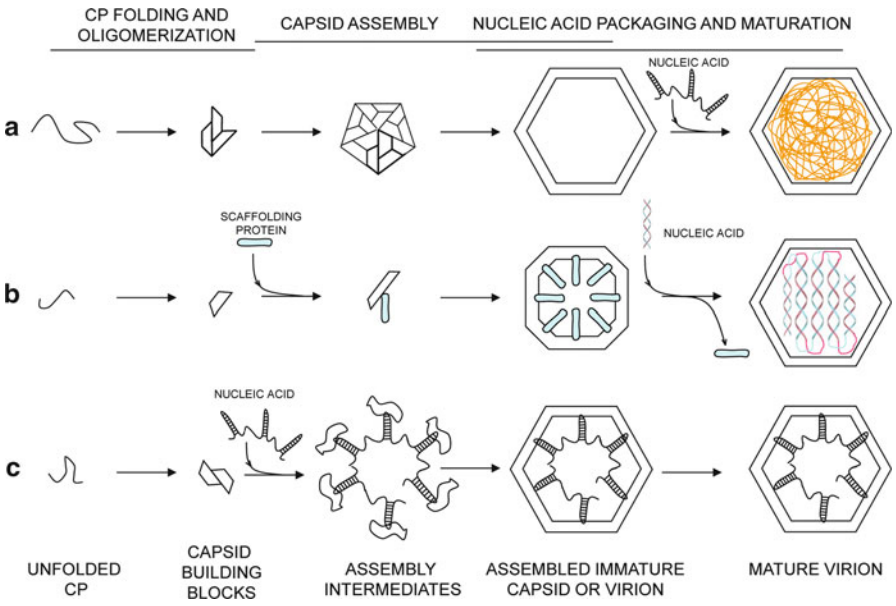


Fig. 1.5 Simplified schemes of three general strategies for assembly of virus capsids. (**a**) Unassisted self-assembly. (**b**) Scaffolding protein-assisted assembly. (**c**) Viral nucleic acid-assisted assembly (see text for details) (Figure kindly provided by M.A. Fuertes and reproduced from [Mateu MG (2013) Arch Biochem Biophys 531, 65–79]. With permission)

Capsid Assembly

Assembly of a virus capsid is a complex oligomerization (some would say multimerization) process that proceeds along a pathway regulated by ordered interactions between the participating macromolecules. The structural complexity of these large oligomers (or multimers) usually requires intricate assembly pathways mediated by allosteric switches, irreversible steps, and the assistance of other molecules including scaffolding proteins or the viral nucleic acid. Three different general strategies for capsid assembly can be recognized (Fig. 1.5): (i) Capsid self-assembly, which strictly requires only CP subunits that self-associate under appropriate conditions (Fig. 1.5a); in some cases, some factors may be required to provide the right conditions for self-assembly. (ii) Scaffolding protein-assisted capsid assembly, which invariably requires not only the CP subunits, but also the assistance of scaffolding proteins to produce a precursor, immature capsid (procapsid) (Fig. 1.5b); (iii) Viral nucleic acid-assisted capsid assembly, which requires the simultaneous interaction of the CP subunits and the viral nucleic acid in a condensation process in which capsid assembly and nucleic acid packaging occur simultaneously (Fig. 1.5c). In this book, and from a morphogenetic perspective only, we have arbitrarily considered as *simple viruses* those that do not require scaffolding proteins for capsid assembly (Chap. 10); and as *complex viruses* those that do require scaffolding proteins and, thus, will follow a more complex capsid assembly pathway (Chap. 11).

Virus capsids are generally assembled from soluble, stable capsid building blocks (CBB) (Fig. 1.5). Depending on the virus, CBBs may be folded CP monomers or, frequently, small oligomers that are formed by association of CPs in a previous step; stable CBBs can be, thus, indistinctly regarded as preformed starting substrates for the capsid assembly reaction itself, or as early intermediates in the complete morphogenetic process from unfolded CPs to native virions (see Chaps. 10, 11). *In vitro* or in the cell, the CBBs may be found in an assembly-incompetent state (depending on virus and conditions), and some environmental change or additional factor may be required to activate them and/or to assist in their association.

Capsid self-assembly. Several important aspects of the thermodynamics and kinetics of capsid self-assembly (Fig. 1.5a) have been experimentally determined using *in vitro* systems; in these systems, simple virus capsids (or virions in a few cases) are assembled from its CBBs under close-to-physiological or, sometimes, non-physiological conditions. Unfortunately, the assembly pathways from CBBs to simple icosahedral capsids have proved to be very difficult to trace experimentally, as populated (stable) intermediates (not counting the initial CBBs) are rarely observed, and transient intermediates can be difficult to detect. Despite these difficulties, scarcely populated, transient assembly intermediates have been identified and characterized during assembly or disassembly of simple capsids, for example using electrospray ionization-MS or ion-mobility-MS. In addition, several stable intermediates have been identified during assembly of a few simple viruses, for example in picornaviruses (see Chap. 10). In contrast, self-assembly and subsequent maturation of more complex icosahedral capsids frequently involve a number of intermediates, including fairly stable intermediates populated at equilibrium and/or transient intermediates, some of which have been identified using different structural or dynamic techniques (see Chaps. 11, 13).

A number of simplified theoretical models and computational simulations on the thermodynamics and/or kinetics of virus capsid self-assembly have been developed (Chap. 19), and contrasted with the results of *in vitro* experiments in which simple virus capsids are self-assembled from their CBBs. In turn, results obtained in these experiments have been used as constraints to develop improved or novel theoretical models. It must be noted that, in general, the theoretical models and simulations are limited to the self-association of CPs or CBBs (although some recent coarse-grained simulations already include the viral nucleic acid). Thus, from a physiological perspective the results thus obtained may be most relevant in those cases in which the capsid self-assembles *in vivo* without the assistance of scaffolding proteins or the viral nucleic acid. However, theoretical and experimental capsid-only approaches may also contribute to (qualitatively) understand the assembly of capsids that, *in vivo*, require scaffolding proteins or the viral nucleic acid. This possibility is supported by the fact that some of these capsids can be correctly assembled *in vitro* without the assistance of any other macromolecules.

A high-order reaction in which 12 pentameric, 20 trimeric or 30 dimeric CBBs collide simultaneously and in the right orientations to form an icosahedral capsid is

unlikely. In fact, theoretical and experimental analyses support the view that simple icosahedral virus capsids can self-assemble from CBBs through a cascade of second-order reactions. Those analyses have also revealed other rather general features: (i) Self-assembly may be thermodynamically considered as two-state: only the dissociated state (soluble CBBs) is populated below a certain CP concentration, and only the fully associated state (complete capsids) is populated above that concentration. (ii) Capsid assembly kinetics can be represented by a sigmoidal curve and includes a lag phase. (iii) The reaction rate is strongly dependent on protein concentration, and this and the previous features suggest that capsid assembly proceeds through nucleation and growth. (iv) At high protein concentrations, free CBBs disappear but only partially assembled capsids are produced. (v) There is hysteresis to dissociation: the capsid disassembles at much lower CP concentrations than those required for assembly. (vi) Off-pathway reactions may occur, leading to aberrant particles, capsids with non-native quaternary structure or polymorphisms. See Chap. 19 for detailed descriptions of theoretical models on capsid self-assembly.

Scaffolding protein-assisted capsid assembly. The proper assembly of many icosahedral virus capsids *in vivo*, and even *in vitro*, requires the assistance of scaffolding proteins (Fig. 1.5b; see Chap. 11). Scaffolding proteins establish specific but transient protein-protein interactions with the CP subunits (CBBs) during the assembly process and are later removed, being absent in the mature virus particle. Thus, scaffolding proteins can be regarded as assembly chaperones. Most, but not all, capsids that require scaffolding proteins for proper assembly have a complex quaternary structure. In some cases the scaffold may self-assemble in the absence of CPs, and is subsequently used as a template for capsid assembly by recruiting CP subunits. In many cases the internal scaffold and the capsid are formed in a co-assembly process where the scaffolding protein subunits promote CP-CP interactions, and the CP subunits promote in turn interactions between the scaffolding protein subunits by eliciting conformational changes in them. Once a precursor, or immature capsid (*procapsid*) is assembled, the scaffolding proteins are removed, either before or during packaging of the viral nucleic acid. In some cases the scaffolding proteins subunits are extruded through openings in the capsid and recycled, thus acting as assembly catalysts. In other cases they are degraded by an encapsidated viral protease.

The use of scaffolding proteins for viral capsid assembly offer additional possibilities for controlling the process. Several functional roles for the scaffolding proteins of different viruses have been identified or suggested: (i) Initiation of capsid assembly by nucleation. (ii) Promotion of capsid assembly by increasing the effective concentration of CPs and/or lowering the energy barrier of a conformational transition that activates the CP or CBB. (iii) Directing assembly; the size of the scaffold would determine capsid size by preventing incorrect CP-CP interactions, and the scaffolding protein-CP interactions would promote the proper conformational switches in the CP subunits being incorporated. (iv) Mediating the

incorporation of other viral structural proteins and subassemblies to the capsid. (v) Preventing the capsid from being filled with some nonspecific proteins which would be difficult to remove. (vi) Stabilization of the assembled procapsid. It must be noted here that other viral proteins or protein assemblies in addition to scaffolding proteins can assist the assembly of some virus capsids, irrespective of its structural simplicity. See Chaps. 10, 11 for detailed descriptions of capsid assembly in different simple or complex viruses.

Nucleic Acid Packaging

Viruses have evolved two general strategies for packaging their nucleic acid genome in the capsid (Fig. 1.5): condensation of the CPs or CBBs and the viral nucleic acid, or insertion of the nucleic acid in a preformed empty capsid (Chap. 12).

Nucleic acid-assisted capsid assembly and genome packaging. In many viruses the viral nucleic acid is recruited to assist the assembly process, to directly yield nucleic acid-containing virions in a combined assembly-packaging process (Fig. 1.5c). In these cases, no empty capsid intermediates are normally formed during virus particle assembly *in vivo*. The experimental observations suggest that, for many ssRNA viruses, the CP subunits (or CBBs) may bind the RNA and influence the acquisition of a defined tertiary structure by the latter, mainly composed of secondary structure elements but not showing a highly compacted, unique fold. In turn, the folded RNA can promote CP oligomerization and influence the geometry of protein-protein interactions between CP subunits to direct the formation of the structurally correct particle. See Chap. 12 for a detailed description of viral capsid and ssRNA co-assembly.

Viral nucleic acid packaging into a preformed capsid. In many other viruses a nucleic-acid-free capsid is assembled first, either without or with the assistance of scaffolding proteins and/or other auxiliary proteins; the viral nucleic acid genome is later packaged into the preformed capsid (Fig. 1.5a, b). This strategy is followed by many dsDNA viruses and dsRNA viruses (see Chap. 12) and also by ssDNA viruses (parvoviruses; see Chap. 10). dsDNA packaging in tailed phages has been extensively studied in exquisite detail. In essence, it occurs as an ATP-driven reaction which involves a molecular motor and a portal protein complex located at one of the vertices of the icosahedron-based capsid, through which the DNA molecule is inserted. See Chap. 12 for detailed descriptions of active dsDNA and dsRNA packaging into preformed viral capsids.

Virus Particle Maturation

Many virus particles are assembled in a non-infectious form. Maturation refers to a generally irreversible reaction, or series of reactions, that convert a non-infectious (immature) virus particle into an infectious (mature) virion (Chap. 13). Maturation reactions may occur before, during and/or after nucleic acid encapsidation. In addition to activation of the virion, a frequent (albeit not universal) effect of

maturation is the physical stabilization of the capsid. Proper assembly of many viral capsids may require weak association energies (see Chap. 19), and the resulting capsids may be thermodynamically unstable. The observed hysteresis to dissociation may be enough, in some cases, to prevent capsid disassembly by extreme dilution during the extracellular phase. However, irreversible post-assembly reactions may have evolved as an additional, or alternative, mechanism to stabilize many virions. Irreversible maturation frequently blocks the capsid in a rather stable (actually metastable, or "spring loaded") state through covalent modification and/or conformational rearrangements, which sometimes involve also the formation of stabilizing covalent bonds and/or binding of additional, "cementing" capsid proteins. Maturation of some other virus particles, for example in retroviruses, may actually destabilize them (see Chap. 18). In considering maturation or any other stage of the viral cycle, it should be taken into account the different selective constraints that may be acting of different viruses depending on their structure and "lifestyle" (*e.g.*, genome protection in enveloped *vs.* non-enveloped viruses, entry of animal viruses *vs.* nucleic acid injection in bacteriophages, etc.).

Probably because of different selective constraints, maturation does take several forms in different virus families and even species. In several families of small, nonenveloped ssRNA viruses, maturation involves a proteolytic cleavage of some capsid subunits. This reaction frequently results in the release of a peptide required for interaction of the virion with cellular membranes during the infection process. In tailed dsDNA phages, maturation overlaps with DNA packaging and is a complex irreversible process that involves capsid conformational changes and achieves different functional goals: removal of the no longer needed scaffolding proteins, capsid expansion, creation of new binding sites for the attachment of additional structural proteins and subassemblies required for virion function, and capsid stabilization. As for simpler icosahedral capsids, the association energy between CPs in tailed phage capsids is low and the procapsid (*prohead*) is rather unstable. The viral dsDNA is packaged into the capsid to a very high (crystalline) density, which results in the capsid becoming pressurized. Capsid pressurization is thought to be required for assisting injection of the viral genome into the infected host bacterium, but will tend to disrupt the unstable procapsid (see Chaps. 9, 17, 18, 19). Thus, different dsDNA phages have evolved different specific strategies for capsid stabilization during and after maturation. See Chap. 13 for a detailed description of these and other maturation strategies and mechanisms in different viruses. Examples of maturation in some other viruses can be found also in a few other chapters in this book (*e.g.*, Chap. 10). Factors contributing to the stability of an already mature virion are briefly described in Sect. 1.4.4.

Host Cell Factors Involved in Virus Morphogenesis

Virion morphogenesis in the infected cell is a very complex process. The cell interior is a molecularly crowded environment, where many competing reactions take place. Macromolecular crowding has been shown to contribute to virus capsid assembly, but it can also promote the formation of aberrant assembly products.

Moreover, replication of the viral genome, capsid assembly, nucleic acid packaging and virus maturation must not only occur, but need to be strictly coordinated in space and time in order to produce an infectious virion. As a result, virion morphogenesis in the host cell generally requires auxiliary viral or cellular factors, including chaperones, and occur in *virus factories*, very complex cellular structures that are formed inside infected cells. Virus factories can be regarded as huge dynamic scaffolds that assist virion replication and assembly; they are currently being studied, for example, by electron microscopy and electron tomography of infected cells. Chapter 14 describes structural, molecular and cell biology techniques that are being used to identify host cell factors that may participate in virus assembly, or for studies on virus morphogenesis in the infected cell; the same chapter describes also some important results obtained on this complex and still poorly known process.

1.4.4 The Virion During the Extracellular Phase

Virus Stability

The many physical and chemical aggressions a virion has to withstand outside cells and organisms have led to multiple structural solutions for virion stabilization; these solutions, however, must not interfere with the capacity of the virion to release its genome into a host cell. Some of the stabilization strategies that different viruses have evolved are summarized below. Studies on the structural determinants of the stability of virus particles are providing further clues on the biological success of viruses; they are also leading to the identification of novel targets for antiviral intervention, and to explore possibilities for the engineering of viral particles of increased stability for bio/nanotechnological applications.

Noncovalent interactions between capsid subunits. Interfaces between CPs or CBBs can differ widely between capsids, or even within a capsid: They can be small or quite large; discontinuous or continuous; quite planar or convoluted; independent of, or overlapping with other CP-CP interfaces; including a hydrophobic core and polar rim, or essentially polar; etc. (see Chap. 10 for a few examples). In general, mutational analyses have identified many specific structural elements and residues in the CPs that are relevant for capsid assembly and/or stability in different viruses. The assembly and/or stability of viral capsids may be critically altered by single amino acid substitutions; in addition to the hydrophobic effect (as estimated by buried surface area), buried electrostatic interactions (hydrogen bonds and charge-charge attractive interactions) may be most important for assembly of a stable capsid in some viruses. Conversely, charge-charge repulsions between CP subunits have been clearly involved in the destabilization of very different virus capsids. Electrostatic repulsions between CP subunits in the capsid can be modulated by protonation/deprotonation both *in vitro* and *in vivo*, and may

contribute to the capsid stability-instability balance and the control of biologically relevant capsid conformational transitions.

Covalent bonds between capsid subunits. Disulfide bonds or other covalent bonds are formed in the capsids of very different viruses; in many cases, they have been shown to transiently or permanently contribute to capsid assembly and/or stability. The reversibility of disulfide bonding (bonds formed in the oxidizing extracellular environment, but likely disrupted in the reducing environment inside the cell) may provide a convenient strategy to meet the conflicting requirements between extracellular stability *versus* intracellular lability for some viruses.

Cementing proteins. The interactions established between scaffolding protein subunits, and between them and the CP subunits, can help to transiently stabilize some immature capsids before they become irreversibly stabilized by maturation. The attachment of bridging or "cementing" proteins during capsid maturation constitutes an evolutionary strategy of some viruses to increase capsid stability by providing additional intersubunit interactions.

Capsid-viral nucleic acid interactions. In many icosahedral ssRNA viruses, the N-terminal and/or C-terminal segments of the CPs are positively charged and normally located in the capsid interior, where they may establish ionic interactions with the negatively charged phosphates of the nucleic acid. In some viruses a fair number of positive residues are located at the capsid inner wall. Other virions contain polyamines. All of these alternatives can neutralize a substantial fraction of the RNA negative charge, and contribute significantly to the nucleic acid-assisted assembly and/or virion stability. In several ssRNA plant and animal viruses and a few ssDNA viruses (parvoviruses), short stretches of the viral nucleic acid molecule conform to the icosahedral symmetry of the capsid and can be directly visualized by X-ray crystallography of the virions. Non-covalent interactions between structured regions at the capsid inner wall and nucleic acid stretches may have a role in RNA-assisted capsid assembly or DNA packaging into a preformed capsid, and may also stabilize the assembled viral particles against a conformational change or dissociation.

Hysteresis to dissociation. For nonenveloped viruses whose capsids are not irreversibly stabilized after assembly through maturation, inbuilt hysteresis to dissociation may be essential to prevent disassembly in the extracellular environment by extreme dilution.

Entropic stabilization. Increased capsid breathing reduces the difference in conformational entropy between the dissociated and associated states. Thus, breathing may help to stabilize the assembled capsid.

Stabilizing ligands. Some ligands (metal ions, antibodies, etc.) may increase the stability of virus particles. For example, small hydrophobic molecules (*pocket factors*) are able to bind hydrophobic pockets in the capsids of some picornaviruses (poliovirus, PV and human rhinovirus, HRV) and can entropically stabilize the virion against conformational changes, inhibiting genome uncoating (see also below).

Covalent attachment of functional groups. Some irreversible enzymatic reactions on viral capsids during maturation stabilize the viral particle, and chemical modifications of specific capsid residues may have different roles in assembly and/or stability.

Mechanical stability of virus particles. Nearly every study on the molecular determinants of viral capsid stability so far has focused on analysis of the resistance of the virion against inactivation of its infectivity (as an indication of structural alterations), or of the viral particle against chemical modification (including proteolytic cleavage), conformational rearrangement or dissociation into subunits by the action of heat or (bio)chemical agents. However, viruses may be also subjected to severe mechanical stress both outside and inside cells, and must be robust enough to withstand the mechanical forces acting on them. Moreover, viruses may have evolved not only to withstand, but also to use those mechanical forces. Very recently, the mechanical stability of virus particles and other aspects of virus mechanics, and their possible biological consequences, have begun to be investigated using AFM (Chap. 18) and physics-based theoretical approaches (Chap. 19).

Virus Dynamics

Virions must achieve enough stability in the extracellular environment without compromising viral genome uncoating in the host cell. This conflict between stability and instability has generally been met by many viruses by evolving a virion that is relatively robust, but conformationally dynamic and metastable.

Breathing. Biochemical and biophysical studies of non-enveloped virions and capsids in solution reveal that that they are not static at equilibrium, but constitute dynamic assemblies which sample a range of similar conformations around a minimum free energy state. This so-called *breathing* tends to transiently increase exposure to solvent of some structural elements in the particle, including functional CP peptide segments, and these events may be biologically required during some steps of the viral cycle. Capsid breathing can be modulated by internal components or capsid-bound external molecules. In the PV and HRV virions a large hydrophobic cavity (pocket) is involved both in breathing, and in facilitating a transition of the particle from a metastable conformation to another conformation during cell receptor recognition and nucleic acid uncoating. Binding of antiviral agents to the hydrophobic pocket (see Chap. 20), or of a virus-neutralising antibody to a region nearby, or mutations in the capsid that may partly fill the pocket were shown to impair viral breathing. The same factors also impair the transition between the metastable and final states of the capsid. The above and other results indicate that global or local conformational dynamics (breathing) may be a general feature in virus particles, and may be required for infection by facilitating structural transitions. Very recent AFM studies have revealed that different regions in viral particles can differ in mechanical stiffness, and indicate that local regions in the particle with high mechanical flexibility may also present high conformational dynamism which may be biologically relevant (see Chap. 18).

Conformational rearrangements in mature virions. Excess stability and rigidity may lead to a biologically inert virion, unable to replicate in a host cell. In addition to conformational dynamism at equilibrium, metastable virions respond to the appropriate stimuli by undergoing controlled conformational transitions required for infectivity. In non-enveloped viruses, the entire viral particle may be conformationally rearranged, although the greatest alterations may sometimes be localized (Chap. 15). In enveloped viruses, a particular structural component (*e.g.*, an envelope glycoprotein on the surface) may change its conformation (Chap. 16). The transitions must be controlled through the action of regulatory factors, to ensure they occur at the appropriate location and time. The same conformational rearrangements caused by physiological factors can be frequently induced *in vitro*, even under non-physiological conditions using different agents (*e.g.*, changes in temperature or pH). These facts have facilitated the *ex vivo* or *in vitro* studies of capsid conformational rearrangements using a vast array of techniques. In some cases, the biological relevance of the transitions observed is not yet clear. In other cases, the identified capsid rearrangements have been shown to be biologically required. These structural changes may be required, for example, for exit of the virion from the cell compartment or the cell where it was formed, entry into another cell, intracellular transport, or uncoating of the viral genome. Specific examples of conformational rearrangements of viral particles or some of their components, and their biological implications, are described in many chapters of this book including Chaps. 10, 11, 12, 13, 15, 16, 17, 18 and 19.

To summarize, the usual end product of virus morphogenesis is a robust enough macromolecular complex able to effectively protect the viral genome during its extracellular existence. However, this complex is also a flexible, dynamic, breathing, metastable particle poised to undergo controlled conformational transitions required to perform biologically critical functions during different stages of the virus life cycle, including entry into cells, intracellular trafficking and viral genome uncoating.

Virion Neutralization by Antibodies and Virus Escape from Neutralization

An extracellular virion must not only be stable enough to withstand physicochemical aggressions without compromising its function; it must also be able to evade the host defenses without compromising virion stability and function. Higher animals, especially mammals, fight viral infection by mounting a complex molecular and cellular immune response (see Chap. 21 for an outline). Antibodies constitute a major part of this immunological mechanism of defense against many viruses. A number of structural studies have focused in the recognition of virions or some of their components by antibodies, the mechanisms by which antibodies neutralize virion infectivity, and the structural strategies viruses follow to escape from antibody recognition. Structural descriptions of complexes between virions or their components and antibodies can be found in Chap. 14 of ref. [15] and Chap. 9 of ref. [17] provides a review on the relationships between the structural basis of virus-antibody recognition and virus escape from neutralization, focusing on foot-and-mouth disease virus (FMDV) as a model system for small, nonenveloped viruses.

From an applied perspective, the information obtained on virus-antibody recognition and virus escape has proven to be critical for the development of improved and novel vaccines including VLPs and chimeric viral particles (Chap. 21)

Virus-antibody recognition. The structures of complexes between individual viral proteins (*e.g.*, envelope proteins of enveloped viruses) and monovalent antigen-binding antibody fragments (Fabs) are generally determined by X-ray crystallography (Chap. 4). However, complexes between non-enveloped virions or capsids and Fabs do not readily crystallize, and combined approaches based in X-ray crystallography of virus particle and Fab separately, and cryo-EM of the virus particle-Fab complex (at lower resolution) are generally used to obtain quasi-atomic models of the ensembles (Chap. 7). In some respects, virus-antibody recognition is structurally similar to virus-receptor recognition and other virus-protein ligand interactions. The reader is referred to Chap. 15 for general structural principles of virus-receptor recognition, some of which can also be applied to virus-antibody interaction (but see below).

In enveloped viruses (such as influenza virus or HIV-1), some antibodies recognise and bind strongly to specific patches (*neutralization epitopes*) on the surface of the envelope proteins. In non-enveloped viruses, some antibodies bind to neutralization epitopes on the capsid surface. In the case of some small non-enveloped viruses (*e.g.*, PV and some types of HRV), a remarkable difference between receptor and antibody recognition is that the receptor-binding site lies in a *canyon*, a depression of the virus surface, while antibodies cannot penetrate very deep into the depression and bind preferentially to capsid protrusions. This is not the case with other similar viruses, including FMDV, whose receptor-binding site is located in a fully exposed, protruding and mobile protein loop (termed the *major antigenic loop*) which also binds many neutralizing antibodies.

On a viral capsid, epitopes may be formed by residues in just one loop of a CP subunit (*continuous epitopes*; *e.g.*, most epitopes in the *major antigenic loop* of some FMDV types), or several segments of a same CP subunit or of several neighboring CP subunits or CBBs (*discontinuous epitopes*; *e.g.*, epitopes identified in other, equally immunodominant antigenic regions in FMDV). If segments from different CBBs are involved, the epitope will be formed only after some intermediate or the complete capsid has been assembled; some will form only after the capsid has been structurally reorganized during virus maturation (Chap. 13). Antibodies against these epitopes can be, thus, used as probes for studying capsid dynamics. It must be noted here that even if the antibody-binding residues are all located in one protein loop, not every residue in a peptide stretch may be directly involved in binding and, strictly speaking, nearly all epitopes are discontinuous. In addition, all viral antigenic epitopes are *conformation-dependent* as, like any other protein ligand, they must adopt a defined conformation either before or, more rarely, on binding the antibody.

About 15–20 residues in the virus capsid or envelope protein may form the *contact epitope*, and interact with a similar number of residues in the antibody *paratope*, contributing to binding strength and specificity through the hydrophobic effect and multiple non-covalent interactions (van der Waals, hydrogen bonds and

charge-charge interactions). However, in many viral epitopes very few residues in the contact epitope, those that form the *energetic epitope*, contribute most of the binding energy. In addition, some residues that may not form a part of the contact epitope may indirectly contribute to binding because, if substituted by mutation, the epitope conformation may be altered, preventing antibody binding. The residues whose mutation affects antibody recognition, located either within the contact epitope or elsewhere in the viral particle, may be considered a part of the *functional epitope*.

Virus neutralization by antibodies. Once bound to a capsid (non-enveloped viruses) or envelope protein (enveloped viruses), antibodies can neutralize virions by one of several different mechanisms. A frequent mechanism is interference with cell receptor or co-receptor recognition. For example, antibodies that bind the major antigenic loop of FMDV overlap extensively with the receptor binding site and sterically interfere with receptor binding (see Chap. 15; [17]). Other mechanisms involve impairment of a conformational rearrangement of a protein or of the virion that would be required at some stage of the viral cycle. For example, some antibodies prevent a conformational transition of the viral fusion protein that is required during entry of enveloped viruses (*e.g.*, HIV-1) into the host cell (see Chap. 16). Some antibodies may exert a neutralizing effect by binding bivalently to the capsid, increasing capsid stability and probably impairing uncoating; they may even act intracellularly after the virus-receptor antibody has been internalized. Other antibodies may exert their neutralizing effect by facilitating virus elimination through *opsonization*. Some mechanisms of antibody-mediated neutralization of viruses have not been elucidated or well characterized yet.

Virus escape from neutralization by antibodies. In general, virus escape occurs through negative selection by circulating antibodies in the organism; it may occur also by genetic drift. RNA virus populations are *quasispecies*, a term that refers to their extreme genetic heterogeneity, being formed by multiple genetic variants in different and sometimes rapidly varying proportions [18]. This generally leads to the frequent emergence of virus variants differing in some phenotypic trait, usually but not necessarily when subjected to the appropriate selection pressure [18]. As a consequence, drug-resistant variants (see Chap. 20) or antibody-resistant variants, especially in the case of RNA viruses (such as FMDV or HIV-1, among many others) will appear most readily; this phenomenon severely hampers viral disease control.

In some RNA viruses, such as influenza A virus, drastic antigenic changes leading to virus escape from antibody recognition may occur abruptly (*antigenic shift*) by genetic reassortment that substitutes at once a glycoprotein type exposed on the viral envelope by an antigenically very different type. This sudden antigenic change poses particularly severe problems for disease control. In addition, in these and many other viruses antigenic variation may occur more gradually (*antigenic drift*), through the accumulation of amino acid substitutions in epitopes located in several antigenic regions of the envelope proteins (in enveloped viruses) or of the viral capsid (in non-enveloped viruses). These amino acid substitutions generally

eliminate some virus-antibody interactions, sterically interfere with the antibody recognition process, and/or distort the epitope conformation.

In PV, some HRV types and other viruses depressions may have been selectively favored as receptor-binding sites (see above) to allow for partial hiding from antibody recognition, thus preserving to some extent receptor binding under antibody pressure (see Chap. 15). However, a clear-cut structural separation between receptor-binding sites (or other functional regions), and antibody-binding sites does not generally occur in small non-enveloped viruses. For example, some antibodies that can neutralize HRV recognise epitopes that include part of the canyon wall and also residues involved in receptor binding. In FMDV the receptor-binding site is exposed and nearly fully overlap with a major antibody-binding site (see above). Thus, additional or alternative mechanisms must operate to preserve cell receptor recognition, and also other virion functions, in the face of a strong antibody selection pressure (see next).

Very thorough genetic mapping of epitopes in small non-enveloped viruses (*e.g.*, FMDV) by selecting many antibody-escape mutants revealed that escape mutations tend to repeatedly occur at very few positions in the contact epitope. These residues may sometimes correspond to those in the energetic epitope, but this is not always the case at all. A paradigmatic case is that of the major antigenic loop of FMDV type C. In this loop, a RGD motif and some neighboring residues are critically involved in binding both the cell receptor and different virus-neutralizing antibodies. FMDV cannot escape from neutralization by any of these antibodies through mutation of nearly any one of the residues critically involved in antibody binding, because these are precisely the ones critically involved also in cell receptor binding. However, the virus manages to readily and repeatedly escape neutralization by adopting either of two non-exclusive alternatives: (i) a drastic one (a lesser kind of antigenic "shift", albeit not due to reassortment, affecting a major antigenic site), by mutation of one of a pair of unique residues that are critical for binding most antibodies that recognise this loop, but less so for binding the receptor; and (ii) a gradual one (a kind of antigenic "drift" in that same antigenic site) by accumulation of mutations of several residues that, individually, are of minor relevance for antibody binding but that, together, severely impair binding of most antibodies that recognise this viral loop. Studies with many FMDV type C field variants isolated during several decades in two continents indicate that antigenic "shifts" in the major antigenic loop involving mutation of those unique residues have occurred repeatedly and independently, but infrequently; this paucity may be related to a significant loss of fitness caused by these "balancing act" mutations. In contrast, antigenic "drifts" in this loop by gradual accumulation of less critical mutations appear to have frequently contributed to FMDV escape in the field.

In small non-enveloped virions, the overlap between antigenic regions and regions involved in virion functions may extend to the entire capsid surface. Competition between antibodies or Fabs for binding to FMDV or other viruses suggest that a large part of the capsid surface is involved in neutralization epitopes. In contrast, viral escape mutations identified in the laboratory or in virus variants isolated in the field collectively tend to cluster in very few discrete regions highly exposed on the capsid surface. Thorough mutational studies with FMDV and other

small viruses indicate that most residues in the capsid surface are involved in virus function: if mutated, they severely impair infection, or at least decrease the biological fitness of the virus. Unlike cellular proteins, the protein capsids of small, nonenveloped viruses may have been functionally *streamlined* through multiple, sometimes conflicting selective pressures imposed on a very simple structure (see Chaps. 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19). As a consequence, only very few residues in the viral capsid can be mutated to evade the antibody response and still allow normal virus infection. Immune evasion of FMDV type C in the field appears to have repeatedly involved mainly combinations of mutations of a surprisingly restricted number of residues in the capsid surface; these capsid residues may be among the few whose mutation has no severe effect on virus function.

1.4.5 Virus-Host Cell Recognition, Entry and Uncoating

After an often long and dangerous voyage in the extracellular environment, in which a few virions may have suceeded in preserving their integrity and functionality while evading the host defenses, one of the surviving virions may eventually encounter another host cell. The virion may then infect the cell by anchoring to appropriate receptor molecule(s), which will trigger the penetration of the entire virion or some of its components (invariably including the viral nucleic acid) into the cell. Unless the nucleic acid enters alone and naked, viral genome replication and expression will require uncoating of the internalized viral nucleic acid. Cell recognition, entry and uncoating frequently overlap and rely on quite different mechanisms depending on the virus species and host type, and are very briefly summarized here separately for animal viruses (see Chaps. 15, 16), bacteriophages (see Chap. 17) and plant viruses.

Cell Recognition, Entry and Uncoating of Animal Viruses

Cell receptor recognition. Like any other cell, animal cells are surrounded by a soft, lipidic plasma membrane. However, unlike bacteria or plant cells, animal cells are not surrounded by a hard, thick cell wall, and this absence facilitates virus penetration. In general, animal viruses start the infection of a host cell by specifically recognizing some *receptor* molecule embedded in the cell membrane (Chap. 15). In the case of non-enveloped viruses, a part of the capsid surface, or a specialized capsid-attached viral protein can bind some cellular molecule (typically a glycoprotein, carbohydrate, or glycolipid) that normally serves some cellular function. In the case of enveloped virions, a viral glycoprotein embedded in the viral lipid envelope can specifically interact with the cell receptor molecule. In some cases,

co-receptor molecules in the cell membrane may also be recruited to complete the recognition process. Virus-receptor binding is an energetically favorable process, and may trigger conformational changes in the bound capsid or viral protein that mediate the entry process.

Entry into the host cell. Attachment of the virus to the receptor molecule starts a complex process in which the entire virus particle, or a part of it (usually the capsid with the nucleic acid inside), enters the cell. Penetration may, in general, occur by direct entry through the cell membrane; or, much more frequently, by *endocytosis* mediated by interaction with the receptor(s), which lead to the engulfment of the virion in an endosome. In this latter case, a subsequent mechanism is required for the release of the viral particle or uncoated genome in the appropriate cellular compartment.

The mechanisms of penetration of non-enveloped animal viruses are in general poorly known, but for some viruses this process is understood in some detail (Chap. 15). For example, cell-receptor recognition by many non-enveloped viruses may activate a process of endocytosis, in which the virion enters the cell largely intact and is transported through one or several different internalization routes, ending with the release of the viral genome in the appropriate cellular compartment. Release of the virus or some component (including the viral genome) in the appropriate compartment may occur after lysis of the endosome and diffusion of the virion or disassembled components (see below); or by permeabilization of the endosome membrane, which may include the transient opening of a channel through which the viral nucleic acid can be translocated.

The mechanism of penetration of enveloped viruses is, in general, better known. Binding of a viral glycoprotein to a cell receptor invariably triggers a conformational change in a viral *fusion protein* that leads to the physical joining (fusion) of the viral envelope and a cellular membrane (Chap. 16). In direct entry (as in HIV-1), after receptor recognition the virus envelope and the plasma membrane are fused. In endosome-mediated entry, after the virus is engulfed in the endosome, acidification leads to the fusion between the virus envelope and the endosome membrane. In both cases, the fusion event allows the diffusion of the viral capsid or nucleocapsid containing the viral genome inside the cell. This is followed by trafficking of the relevant viral component to the appropriate cellular compartment for genome expression and replication.

Genome uncoating. Uncoating is, in general, a very poorly known step in the life cycles of animal viruses. In some non-enveloped virions, such as the picornaviruses PV and HRV, receptor binding and or low pH mediate capsid conformational rearrangements that may open pores or fractures large enough for nucleic acid extrusion. In other non-enveloped virions such as the picornavirus FMDV, acidification at the endosomes dissociates the viral capsid into subunits, and the nucleic acid is released into the cytoplasm. Many details of these processes are still uncertain or unknown (see Chap. 15). In enveloped virions, the viral nucleic acid-containing capsid or nucleocapsid that entered the cell after fusion of the virus and cell membranes is transported to the appropriate cell compartment for viral nucleic acid uncoating. In some cases, the capsid is disassembled and the nucleic

acid just released, either by dilution or by a poorly characterized mechanism that may require the participation of cellular factors (*e.g.*, in HIV-1). In other cases, direct replication and expression of the genome may occur with the viral nucleic acid being still a part of a ribonucleoprotein particle (*e.g.*, in influenza virus) or being still associated with a viral capsid core (*e.g.*, in reoviruses). See Chaps. 15 and 16 for a detailed description of different mechanisms of cell recognition, entry and uncoating in different non-enveloped or enveloped animal viruses.

Cell Recognition and Nucleic Acid Transfer by Bacteriophages

Bacterial cells are surrounded by thick and hard cell walls that bacteriophages cannot penetrate. Like animal viruses, bacteriophages recognize cellular receptors on the cell surface. However, the phage particle does not actually enter the cell. The barrier posed by the cell wall is overcome in different ways. In many dsDNA phages, a tube penetrates the cell wall as if it were the needle of a hypodermic syringe, and the nucleic acid genome is injected inside the cell, driven in part by the internal pressure built-in in the capsid during packaging of the nucleic acid. See Chap. 17 for a detailed description of cell receptor recognition by different bacteriophages and phage nucleic acid delivery into the host bacterium.

Entry of Plant Viruses

Plant cells are surrounded, like bacterial cells, by a rigid cell wall a virus cannot easily penetrate. Nearly all plant viruses are ssRNA(+) viruses with a slender helical capsid or a relatively small icosahedral capsid. Most seem to enter the plant cell in a passive way, through breaches in the cell wall and transient openings in the plasma membrane that may be caused by mechanical injury, for example by the action of herbivorous animals or insects. Once a cell is infected, progeny virions may propagate from cell to cell in the organism through discontinuities in the cell wall crossed by channels that establish interconnections between neighboring cells (*e.g.*, plasmodesmata). A general description of plant virus entry and infection can be found, for example, in the book by Hull [8].

1.5 Experimental and Theoretical Developments in Physical Virology

Physical virology (see Sect. 1.1.2 for a brief introduction to the field) is becoming integrated with classical structural virology and providing novel fundamental insights into the structure, dynamics and properties of virus particles. In this book we have described in detail dominant experimental (Chaps. 8, 9) and theoretical (Chap. 19) methods and several basic studies (Chaps. 18, 19) in Physical Virology, as well as some potential applications (Chap. 22).

Two very important physics-based, single-molecule techniques that have started to be applied to viruses are AFM (Chap. 8) and optical tweezers (Chap. 9). AFM can be used to image viruses and follow structural changes in viral particles in real-time (with the temporal limitations imposed by the time required to acquire each image). In addition, AFM can be used to probe some physical properties of virus particles. In particular, their mechanical characteristics such as particle stiffness, Young's modulus of the capsid material (as a measure of its intrinsic stiffness), tensile strength or resistance to mechanical fatigue are being determined. A dominant method used to study virus mechanics involves the indentation of virions or capsids by "pushing" on defined points on the particle surface with the AFM tip (Chap. 18). AFM is also being used to determine the values of the physical forces involved in molecular recognition between virus particles and different ligands, for example cell receptors or antibodies, even in physiologically relevant conditions, with receptors embedded in the membrane of a cell or liposome. The general approach in these studies involves attaching a virus particle or virus ligand to the AFM tip and the interacting partner to a solid substrate, and "pulling" until the interaction between them is broken. Mechanical aspects of some stages in the viral cycle, such as virus entry into cells, can also be followed by AFM (see Chap. 8). Optical tweezers have been particularly successful in determining physical forces involved in the action of molecular motors during the mechanochemical process of dsDNA encapsidation inside a preformed bacteriophage capsid (see Chap. 9).

Several theoretical physics-based approaches are being applied to understand the physical foundations of virus capsid architecture; material properties of viruses, especially virus mechanics; and different aspects of the infection process in which viral particles are involved, including capsid assembly and maturation and virus entry into cells. In this book we have integrated theoretical studies in physical virology in a single monographic chapter (Chap. 19). In that chapter, a theoretical physics-based perspective is used to contemplate many aspects of virus structure, dynamics and properties that had been described in previous chapters (Chaps. 2, 10, 15 and 18) from an empiric, structural biology-based perspective.

1.6 Applied Structural and Physical Virology

Several decades of intensive research on the structure, dynamics and properties of virus particles, part of which is described in Chaps. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19 of this book, have opened up many possibilities for the exploitation of the knowledge acquired, and of the viral particles themselves, in biomedicine and biotechnology. In addition, the recent advent of nanotechnology and novel studies on the physics of viruses has led to a growing awareness by researchers of physical or chemical disciplines on the multiple possibilities virus particles present also for nanotechnological applications. In addition to their possible direct use, viruses are also providing inspiration for the design of non-viral or

hybrid nanoparticles and nanomaterials. Among others, two extremely important applied areas in human and animal health in which the knowledge acquired on virus structure is having a clear impact are the development of novel antiviral drugs (Chap. 20) and virus vaccines (Chap. 21). Knowledge on the structure and physical properties of virus particles is also being applied for the development of viral-based nanoparticles, nanomaterials and nanodevices for multiple applications (Chap. 22)

1.6.1 Virus Structure-Based Antiviral Research

The knowledge acquired from structural studies on virus morphogenesis (Chaps. 10, 11, 12, 13 and 14) and cell receptor recognition and entry of animal viruses into cells (Chaps. 15 and 16) have provided excellent opportunities to develop new antiviral agents based on the inhibition or misdirection of these processes. Chapter 20 focuses on the general structural basis of antiviral action and describes principles and methods for the structure-based search and design of new antiviral agents. Of particular relevance for the subject of this book, the chapter includes several examples where the target of antiviral research is virus assembly, cell receptor recognition or entry into the host cell.

1.6.2 Applications of Virus Particles

Virus particles have acquired through evolution outstanding properties and functions. To name just a few: capsid self-assembly; capsid metastability and conformational rearrangements; the mechanochemical action of viral molecular motors for nucleic acid packaging; the targeting of specific cells through precise molecular recognition; and chemical and mechanical actions for delivery of the genome into the host cells, including particle penetration and controlled particle disassembly, or nucleic acid injection. These properties have opened up many possibilities to turn viral particles into useful biomaterials and devices in biomedicine, biotechnology and nanotechnology. However, viruses as evolved in nature do not have all of the several properties and functionalities required, or are not optimized, for any of their contemplated applications. Thus, methodologies have to be devised to manipulate virus particles and modify their structure, properties, and/or functions to serve the intended goals. Virus particles can be modified by direct chemical means and/or by protein engineering approaches using genetic techniques (see Chap. 22). Learning to successfully modify virus particles may not only provide useful products, but also lead to a deeper understanding on the structure and function of these and other biomolecular complexes.

A non-comprehensive list of current or under-development applications of the knowledge acquired on virus particles, and of adequately engineered viral particles may include (Fig. 1.6):

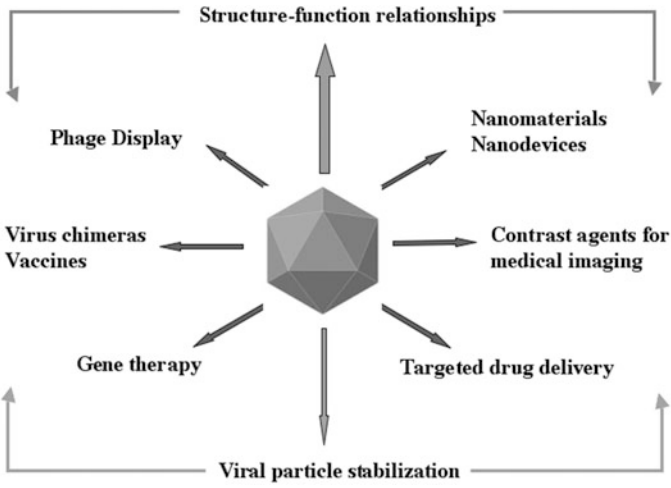


Fig. 1.6 Some applications of virus particles in biomedicine, biotechnology and nanotechnology (see text for an explanation)

(i) Further structure-function studies of viruses oriented to explore, evaluate or implement novel applications.
(ii) *Phage display* techniques for the selection from a combinatorial library of variant proteins or peptides of those with the desired properties (binding to specific ligands, improved binding, chemical or conformational stability, catalysis of specific substrates or improved catalytic efficiency, etc.).
(iii) Novel or improved vaccines, including those based on virus-like particles (VLPs) obtained by production of CPs in expression systems and self-assembly in the cell or *in vitro*, or chimeric VLPs or virions carrying foreign epitopes from a pathogen (Chap. 21).
(iv) Delivery of therapeutic genes into specific cells for gene therapy.
(v) Targeted delivery of drugs using viral particles as nanocarriers or nanocontainers for specific chemotherapies.
(vi) Target-specific contrast agents for molecular imaging and diagnosis.
(vii) Building blocks for the construction of nanomaterials, nanostructures and nanodevices for technological uses, from medicine to electronics. Chapter 22 describes a number of viral-particle-based possibilities being explored for applications in biotechnology and nanotechnology.

1.7 Concluding Remarks

A vast amount of information on the structure, dynamics and properties of virions and their molecular components has accumulated at an ever increasing pace. The interplay between structural, physical and molecular virology has led to a giant leap

in our understanding of virus structure and function; a deeper and better comprehension on the workings of the molecular machinery of life; critical advances in the fight against viral disease; and promising developments in the application of virus particles to medicine and technology. However, there is still much to be learned on how viruses are made and how they work.

I would like to end this chapter with a note intended for the young student: Truly inspiring views of coral reefs are obtained not by snorkeling, but by diving. William Beebe recommended readers of his 1928 book *Beneath Tropic Seas* (Putnam's sons, NY): "Don't die without having borrowed, stolen, purchased or made a helmet of sorts, to glimpse for yourself this new world". This chapter has provided just a quick look from the *surface* at some of what has been learned on virus structure and physics so far. By *diving deep* into the chapters that follow, the reader will be rewarded with amazing close-up views of these fascinating molecular machines.

Acknowledgements I gratefully acknowledge former and current collaborators and members of my group for their invaluable contributions to our studies on virus structure-function relationships, Miguel Angel Fuertes for providing figures and technical assistance, and Pedro J. de Pablo, Julio Gómez-Herrero, David Reguera and Ricardo Arias-Gonzalez for fruitful discussions on the connection between the physics and biology of viruses. Work in the author's laboratory is funded by grants from the Spanish Government (BIO2009-10092 and BIO2012-37649) and Comunidad de Madrid (CM) (S-2009/MAT/1467), and by an institutional grant from Fundación Ramón Areces. The author is an associate member of the Institute for Biocomputation and Physics of Complex Systems, Zaragoza, Spain.

References and Further Reading[1]

Molecular Virology Textbooks[2]

1. Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2009) Principles of virology, 3rd edn. ASM Press, Washington, DC. *Intermediate level, highly recommended book on molecular virology*
2. Dimmock N, Easton A, Leppard K (2007) Introduction to modern virology, 6th edn. Blackwell, Malden. *Introductory text on general principles in molecular virology*
3. Carter JB, Saunders VA (2007) Virology. Principles and applications. Wiley, Chichester. *Introductory text on general principles in molecular virology*

[1] *Because of the elementary, student-oriented nature of this introductory chapter, we have purposefully limited the list of references to some recommended books. For further references to scientific reviews and original articles on structural and physical virology, the reader is referred to the sections on "References and further reading" included in the rest of the chapters in this book.*

[2] These books contain elementary to intermediate information on the molecular biology of viruses, with emphasis on viral genome replication and expression. Reference to some structural aspects directly related with these processes, which are out of the scope of the present book, may be also found in these books.

4. Wagner EK, Hewlett MJ, Bloom DC, Camerini D (2008) Basic virology. Blackwell, Malden. *Introductory text on general principles in molecular virology*
5. Cann AJ (2012) Principles of molecular virology, 5th edn. Academic Press, Waltham. *Introductory text on general principles in molecular virology*
6. Acheson NH (2007) Fundamentals of molecular virology, 2nd edn. Wiley, Hoboken. *Introductory text dealing separately with the molecular biology of many virus families*
7. Kutter E (2005) Bacteriophage: molecular biology and applications. CRC Press, Boca Raton. *A molecular virology book specialized in bacteriophages*
8. Hull R (2009) Comparative plant virology, 2nd edn. Academic Press, Burlington. *A specialized book on the molecular biology of plant viruses, including many structural features*

Advanced Reference Books on Virology

9. Knipe DM, Howley PM (eds) (2007) Fields virology, 5th edn. Lippincott Williams & Wilkins, Philadelphia. *Advanced, extensive coverage of virology; a classic two-volume reference book*
10. Mahy BWJ, van Regenmortel MHV (eds) (2008) Encyclopedia of virology, 3rd edn. Academic Press, Boston. *Advanced, extensive coverage of virology topics based on concise reviews; a five-volume reference book*

Advanced Books on Structural or Physical Virology[3]

11. Casjens S (ed) (1985) Virus structure and assembly. Jones and Bartlett, Boston. *Largely outdated, but very useful for understanding some fundamental principles in structural virology*
12. Chiu W, Burnett RM, Garcea RL (eds) (1997) Structural biology of viruses. Oxford University Press, New York. *Outdated in many respects, but an extremely useful book that describes clearly many principles in structural virology and contains detailed advanced monographs on a variety of structural virology topics*
13. Chiu W, Johnson JE (eds) (2003) Virus structure. Advances in protein chemistry, vol 64. Academic Press, New York. *An excellent collection of advanced reviews on a limited but highly relevant selection of structural virology topics*
14. Stockley PG, Twarock R (eds) (2010) Emerging topics in physical virology. Imperial College Press, London. *An excellent and largely up-to-date collection of advanced reviews on a number of specific topics in theoretical or experimental physical virology. Highly recommended*
15. Agbandje-McKenna M, McKenna R (eds) (2011) Structural virology. RSC Publishing, Cambridge. *An excellent and up-to-date book containing monographs on many structural virology topics. Some chapters in that book complement others in the present book by covering topics such as evolution of viral capsid structures, structural aspects of virus-antibody recognition, or viral vectors for gene therapy. Physical virology-specific techniques (AFM, optical tweezers, physics-based models and simulations) and studies using these approaches are not contemplated. Highly recommended*

[3] These specialized books contain excellent advanced monographs on many structural or physical virology topics.

16. Rossmann MG, Rao VB (eds) (2012) Viral molecular machines. Advances in experimental medicine and biology, vol 726. Springer, New York. *A collection of excellent and up-to-date, quite advanced and detailed reviews on some structural and molecular aspects of viruses considered as molecular machines, with emphasis on nucleic acid packaging. Physical virology-specific topics (see comment on ref. 15) are not contemplated. Highly recommended for advanced, specialised reading*

References on Virus Genetic and Phenotypic Variation and Evolution[4]

17. Sobrino F, Domingo E (eds) (2004) Foot and mouth disease: current perspectives. Horizon Bioscience, Wymondham/Norfolk. *See Chaps. 10 and 9, respectively, for detailed accounts on quasispecies and virus evolution and antibody recognition and viral escape, using a thoroughly studied virus model, FMDV*
18. Domingo E, Biebricher C, Eigen M, Holland J (2001) Quasispecies and RNA virus evolution: principles and consequences. Landes Bioscience, Austin. *An excellent book on RNA virus variation, quasispecies and evolution*

[4] Important to understand the evolutionary concepts and genetic mechanisms behind the origin and continuous modification and adaptation of the viral machines described in the present book.

Chapter 2
The Basic Architecture of Viruses

José R. Castón and José L. Carrascosa

Abstract Viruses are elegant macromolecular assemblies and constitute a paradigm of the economy of genomic resources; they must use simple general principles and a very limited number of viral components to complete their life cycles successfully. Viruses need only one or a few different capsid structural subunits to build an infectious particle, which is made possible because of two reasons: extensive use of symmetry and built-in conformational flexibility. Although viruses from the numerous virus families come in many shapes and sizes, two major types of symmetric assemblies are found: icosahedral and helical particles. The enormous diversity of virus structures might be derived from one or a limited number of basic schemes that has become more complex by consecutive incorporation of structural elements. The intrinsic structural polymorphism of the viral proteins and other observations indicate that capsids are dynamic structures. Study of virus structures is required to understand structure-function relationships in viruses, including those related to morphogenesis and antigenicity. These structural foundations can be extended to other macromolecular complexes that control many fundamental processes in biology.

Keywords Capsid • Nucleocapsid • Helical symmetry • Icosahedral symmetry • Triangulation number • Conformational polymorphism • Quasi-equivalence • Prolate capsid • Metastable capsid • Molecular switch

Abbreviations

3D	Three-dimensional
CP	Capsid protein

J.R. Castón (✉) • J.L. Carrascosa
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jrcaston@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_2,

cryo-EM	Cryo-electron microscopy
cryo-ET	Cryo-electron tomography
ds	Double-stranded
ss	Single-stranded
T	Triangulation number

2.1 Introduction

Viruses are probably the most extreme examples of nature's efficient use of limited coding capacity. Optimization of resources is reflected in their multiplication and dissemination strategies, as well as in their construction. Viruses use simple, general principles to successfully complete their life cycles [1]. They need only one or a few proteins to build a protein shell (the capsid) and package the viral genome inside to yield an infectious particle (the virion). Virions are endowed with a remarkable number of essential functions, including protection of the genome during extracellular transport, entry into the host cell, intracellular trafficking, genome uncoating and, in some instances, viral genome replication. All of these activities are carried out efficiently, even in the face of structural variability of the capsid proteins (CP) due to distinct selective pressures.

Optimization of resources in building a viral particle has been achieved in two ways: extensive use of symmetry and conformational polymorphism of the CPs [2], both of which allow the virus to build a three-dimensional (3D) container of defined size and shape by using many copies of only one or a few proteins. Except in some complex viruses, the two main classes of symmetric capsid structures are icosahedral and helical [3, 4]. In both cases, the need for multiple copies of one or more CP subunit type arranged symmetrically can also be inferred from the nature of the genetic code. A base triplet (codon) has an approximate molecular weight of 1,000 Da and encodes a single amino acid with an average molecular weight of 150 Da. At best, a nucleic acid can encode 15 % of the virus weight as protein; nevertheless, 50–90 % of the virus molecular weight is protein. This high protein: nucleic acid ratio is necessary to allow complete encapsidation of the genome in a large enough, hollow container through CP oligomerization.

The classical vision of viral capsids as inert containers that protect the fragile genome during the extracellular phase of the virion is inappropriate; capsids are dynamic structures whose components have transient conformations related to specific functions in the viral cycle [5, 6]. The capsid is a metastable assembly, and should be considered a compromise: it is robust enough to protect the genome and ensure its propagation during passage from one host cell (or organism) to another, and labile enough to allow genome delivery in the host cell to initiate infection.

CP subunits (a single folded polypeptide chain) may be grouped in protomers (structural units or "capsid building blocks") from which a capsid is built.

Protomers may be identified by biochemical techniques, and may comprise one or more nonidentical protein subunits (*e.g.*, VP1, VP2, VP3 and VP4 in poliovirus, the smallest subset of CP subunits that may reconstitute the complete capsid through symmetry operations). Protomers may cluster into capsomers, morphological units such as hexamers, pentamers, trimers or dimers of protomers. Capsomers correspond to the apparent clusters, knobs or projections seen on the particle surface by electron microscopy. Because capsomers do not necessarily correspond to chemical entities that can be isolated, this term is generally restricted to descriptions of viruses from electron micrographs.

Interactions among protein subunits in the viral capsid are noncovalent and generally weak. There are exceptions, such as the bacteriophage HK97 protein shell, in which pentameric and hexameric capsomers are covalently joined to form rings that are also topologically interlocked, creating a protein chainmail [7]. For the most part, however, as a result of viral particle symmetry, numerous weak interactions (hydrogen bonds, salt bridges, van der Waals forces and hydrophobic interactions) must cooperate to maintain an appropriate 3D structure and to facilitate dynamic structural processes.

How is a closed shell with a defined shape built? This is a complex process that must reconcile two apparently contradictory requisites; energy requirements must be minimal to ensure high assembly efficiency, which decreases the occurrence of aberrant assemblies *in vivo*, and the protein container should be stable and robust. A number of factors increase the effectiveness of assembly. Parts of the tertiary structure of the CP have built-in conformational flexibility, which allows them to maintain the slightly different intersubunit contacts necessary for the assembly of most icosahedral capsids (see below). CPs have also large interaction areas and, even isolated from their natural environment, CP subunits or small CP oligomers (*i.e.*, capsid building blocks) are able to self-assemble spontaneously *in vitro* into complex assemblies similar or related to viral particles (viral-like particles or VLPs; see Chap. 21). This spontaneous assembly indicates that the viral particle is in a free-energy minimum state. To facilitate and direct correct intersubunit interactions, viruses can use host resources such as molecular chaperones and membrane-specific regions, or other viral components such as scaffolding, accessory, cement, and proteolytic proteins.

In this chapter, we describe the principles of the basic structural organization of viruses and discuss numerous examples of capsid architecture. X-ray crystallography and 3D cryo-electron microscopy (cryo-EM) techniques have shown the molecular mechanisms by which identical protein building blocks make use of symmetry and structural polymorphism to assemble a multifunctional container in a biologically feasible period of time.

2.2 How Virus Structures Are Studied

Viruses were first visualized by electron microscopy, which provided the basis for their classification following morphological criteria. Development of electron microscopy techniques has had a central role in the study of viral structure and

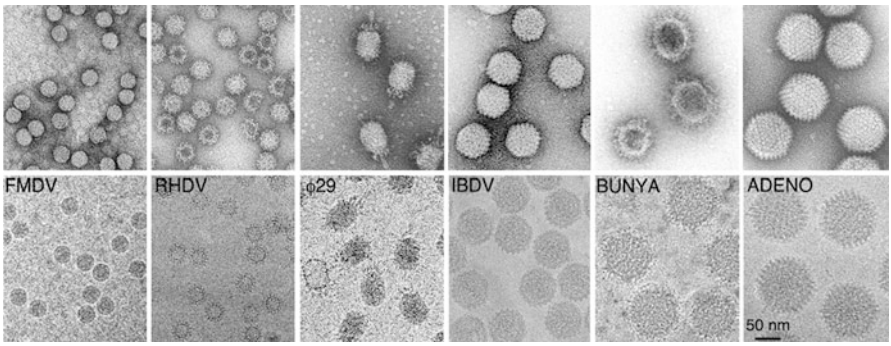


Fig. 2.1 Conventional EM and cryo-EM analysis of several viruses. A collection of representative virus preparations visualized by conventional negative staining (*top*), or directly vitrified without staining and/or fixing agent (*bottom*). FMDV, foot-and-mouth disease virus (sample provided by E. Domingo, CSIC); RHDV, rabbit hemorrhagic disease virus (provided by J. Bárcena, INIA); ϕ29, prolate bacteriophage ϕ29; IBDV, infectious bursal disease virus; BUNYA, the enveloped bunyavirus, Bunyamwera virus (provided by C. Risco, CSIC); ADENO, adenovirus (micrographs by C. San Martín, CSIC)

life cycle (see Chap. 3). The introduction of adequate fixation methods (chemical or physical) and contrast protocols (such as negative staining) for viral samples have allowed observation of isolated viruses and virus-infected cells in thin sections. More recently, cryo-EM has become a major tool; samples are frozen so rapidly that they are immobilized in an amorphous (or vitreous) solid state, without physical damage of the specimen by ice crystal growth [5]. Cryo-EM allows direct visualization of viral particles, as no staining or fixative agents are introduced during sample preparation, and native conformation is preserved, as samples remain hydrated (Fig. 2.1). Combined with image reconstruction techniques, cryo-EM is used widely to establish the 3D structure of viruses. Cryo-EM has contributed notably to demonstration of the metastable nature of viral capsids thoughout their life cycle. Dynamic studies of viruses have shown conformational changes of structural subunits that helped to explain the underlying molecular mechanisms of specific processes. 3D cryo-EM maps are usually calculated at subnanometer resolution, enabling us to infer secondary structure elements in the capsid structural subunits. A few 3D cryo-EM maps have recently been calculated at near-atomic resolution, to trace amino acid backbones (Cα model building) and to define the surface contacts among structural components of the virion [8, 9]. Enveloped virions are flexible (pleomorphic) and less symmetric than naked icosahedral or helical viruses; cryo-electron tomography (cryo-ET) is used to study individual particles of complex enveloped viruses such as orthomyxo- and poxvirus, albeit at much lower resolution than single particle cryo-EM. Cryo-ET is used also to study structural aspects of the viral life cycle directly in infected cells (see Chap. 3).

X-ray diffraction techniques are the other major approach for determining virus structure and interactions at the atomic level. Their major limitation is the need to procure sufficiently large (milligram) quantities of highly pure (and stable) virus

preparations to form suitable 3D crystals for X-ray diffraction. These crystals are relatively easy to obtain from small spherical virus particles, although resolution of complex icosahedral capsids is becoming reasonably common, due to the combination of robotized crystallization procedures and the use of improved X-ray diffraction equipment allowing the structural resolution from very small microcrystals (see Chap. 4). The first high resolution virus structure, that of the ssRNA plant virus tomato bushy stunt virus (TBSV), was determined in 1978 [10]. The first structures of small human viruses, poliovirus [11] and rhinovirus [12], were obtained in the mid-1980s. Atomic resolution maps of viral capsids have been essential in establishing the mechanisms that govern assembly and stability of viral structures and the structural basis of many virus functions.

Despite their minimal sequence similarity, many CP of spherical viruses (in animals and plants) adopt a very limited number of tertiary folds, the most common of which is the antiparallel β-barrel [13]. The β-barrel fold is also known as a jelly-roll or a Swiss-roll β-barrel. This wedge-shaped structure is formed by eight antiparallel β-strands (with two opposing BIDG and CHEF β-sheets) and two α-helices (Fig. 2.2a). CP size varies between 20 and over 70 kDa, with the major differences located in N- and C-terminal ends (usually facing the interior and the outer surface of the capsid, respectively) and in the loop sizes between β-strands (in some cases, more than 100 amino acids that form additional domains). The β-strands usually lie roughly tangential to the capsid surface, defining a protein shell ~30 Å thick, although they can also lie orthogonally relative to the particle surface. The wedge-like shape allows the subunits to pack tightly and, due in part to this built-in flexibility, the jelly-roll motif-based CP can assemble into unrelated capsids of variable sizes (by increasing the T number, see below). The β-barrel is also found in some bacilliform viruses such as alfalfa mosaic virus (AMV).

β-barrel duplications have also been reported. Joint folds were first observed in the adenovirus trimeric capsomer; each monomer consists of two successive jelly-roll motifs, producing a pseudo-hexameric structure [14]. Many other large double-stranded (ds) DNA viruses have similar trimeric capsomers, such as the *Paramecium bursaria chlorella* virus 1 (PBCV1) [15], the dsDNA bacteriophage PRD1 [16], and the sulfolobus-infecting virus, STIV (sulfolobus turreted icosahedral virus) [17]. The large subunit of comoviruses, single-stranded (ss) RNA viruses that infect plants, is also made by fusion of two β-barrel domains [18]. A similar fused fold is found in a transient structural protein in a nonicosahedral virus, vaccinia virus [19].

Crystallographic analysis showed that, in addition to the β-barrel as the most extended fold among CP of spherical viruses [13], a limited number of other folds is used to form viral capsids, including those found in RNA bacteriophage MS2 or in dsDNA icosahedral bacteriophages such as the HK97 phage [20]. The HK97 CP forms an L-shaped, mixed α/β structure with two domains. The peripheral (P) domain consists of a long α-helix and an elongated, three-stranded β-sheet plus extensions (N-arm and E-loop) and the axial (A) domain, composed of two α-helices and an additional β-sheet. The two residues that form covalent crosslinks with neighboring subunits are on opposite sides of the protein (Fig. 2.2b). This structure is optimal for CP self-assembly to form a metastable particle, in which subunit interactions are

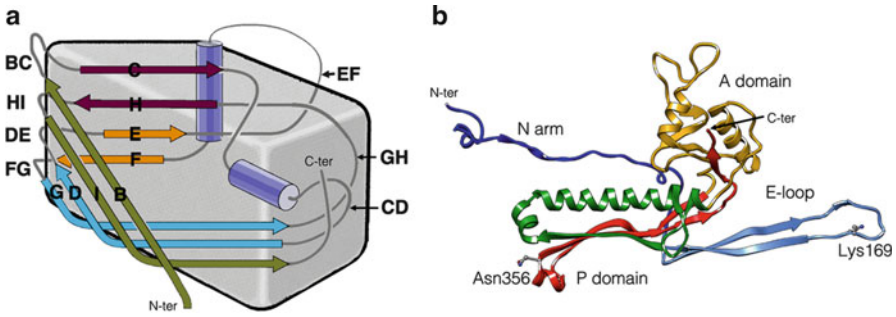


Fig. 2.2 The wedge-shaped β-barrel fold and the HK97 capsid protein fold. (**a**) Scheme showing the connecting loops between secondary structural elements (eight β-strands, arrows labeled B-I, and two α-helices, shown as cylinders). Loops CD, EF and GH generally have large insertions; loops BC, HI, DE and FG are short. This wedge-like domain has large interaction surfaces, appropriate for icosahedral symmetry. (**b**) Structure of the HK97 capsid protein (gp5), color-coded by domains (PDB entry 1OHG). The two domains A and P are indicated, as well as the extended N-arm (*dark blue*) and the E-loop (*cyan*). In the mature capsid, Lys^{169} on the E-loop forms a covalent bond with Asn^{356} on a neighboring subunit

modulated by the CP N-terminal scaffolding domain. A proteolytic process then takes place to remove the scaffolding domain, followed by massive conformational changes and crosslinking, leading to a robust mature capsid [20] (see Chaps. 11 and 13). Some bacteriophages incorporate additional domains in the main shell protein fold, resulting in different assembly mechanisms [21].

X-ray crystallography is used not only to solve atomic structures of entire viral capsids, but also to obtain detailed structural maps of isolated viral proteins such as hemagglutinin and neuraminidase of influenza virus, several viral proteases (from picornaviruses, herpesviruses, hepatitis B virus (HBV), human immunodeficiency virus (HIV), etc.), the HIV reverse transcriptase and several RNA polymerases, among many others. Large viruses and intermediate states of capsid assembly are usually refractory to high-resolution analysis by X-ray crystallography, but are accessible by 3D cryo-EM. Many studies have successfully combined these complementary techniques; pseudoatomic models of the capsids are derived after docking atomic structures of individual capsid components into lower resolution cryo-EM density maps (see Chap. 7). Nuclear magnetic resonance spectroscopy has also contributed valuable structural information by revealing the structure of viral protein domains and the structural dynamics of viral protein domains and capsids (see Chap. 5). Atomic structures of virions, capsids and viral proteins have supplied valuable information for the design of diverse antiviral drugs, including some that alter viral particle stability, interfere with capsid assembly or impair viral enzyme activity (see Chap. 20); they have been also extremely useful for the engineering of virus particles for biomedical, biotechnological or nanotechnological applications (see Chaps. 21 and 22).

2.3 Viral Capsid Symmetry

Molecular crystals are formed from multiple copies of single molecules that establish identical interactions with adjacent molecules, and thus have equivalent environments in the crystal. This is the simplest solution to form a regular or symmetric structure. There are fundamental differences between crystals and viruses, however. Whereas a crystal is a continuous 3D network with no size limitations, viruses build an empty container of defined shape for nucleic acid packaging in a biologically feasible time frame [22]. The capsid shell is therefore seen as a two-dimensional crystal that is limited in size and, in most cases, closed, leaving a cargo space inside. Although they use basic principles of crystal growth, viruses require additional mechanisms to control their dimensions [23].

There are several alternatives for generating 2D crystals within the 17 planar crystallographic groups. Some of these groups are variants of hexagonal lattices. In a hexagonal network, identical subunits can be arranged to maintain them in identical environments. If correct distortions are introduced, two types of closed shells can be made from this planar network, helices and icosahedra. If after bending and binding, the network edges are out of register, a tube is obtained in which layer lines are in register and generate a helix; if the edges are in register, a cylinder with a pile of stacked discs is obtained (Fig. 2.3a). In a helix, structural subunits have identical interactions, except those located at the ends, and the self-assembly of protomers can easily be envisaged. The diameter of the tubular structure depends on the curvature in the subunit interactions, whereas its length is limited by that of the nucleic acid it encloses and by subunit-nucleic acid interactions.

The icosahedron is a closed polyhedron with cubic symmetry, and asymmetric units can therefore be laid on the symmetric polyhedron surface with identical interactions among them. In 1956, Watson and Crick proposed the first principles for construction of spherical viruses [24]. In theory, a limited number of regular solids (termed Platonic polyhedra) allow close packing of repeated (asymmetric) subunits such as proteins. These polyhedra are the tetrahedron (four triangular faces and 12 equivalent positions), the cube (six square faces and 24 equivalent positions), the octahedron (eight triangular faces and 24 equivalent positions), the dodecahedron (12 pentagonal faces and 60 equivalent positions), and the icosahedron (20 triangular faces and 60 equivalent positions). The icosahedron (and the dodecahedron) is the most economic natural structure in terms of genetics, as a larger number of identical subunits encoded by a single gene can be laid in identical positions. In addition, the icosahedron is the cubic symmetrical structure in which the surface/volume ratio is lowest, involving the best component use to generate a hollow polyhedron, that is, a container, of the highest possible capacity per unit of genetic information required to encode it. An icosahedron can be generated from a hexagonal planar lattice by conversion of planar hexagonal nodes (or hexamers) at defined positions in the lattice into convex pentagonal nodes (or pentamers) after removing one subunit/hexagonal node (Fig. 2.3b). Inclusion of pentamers allows

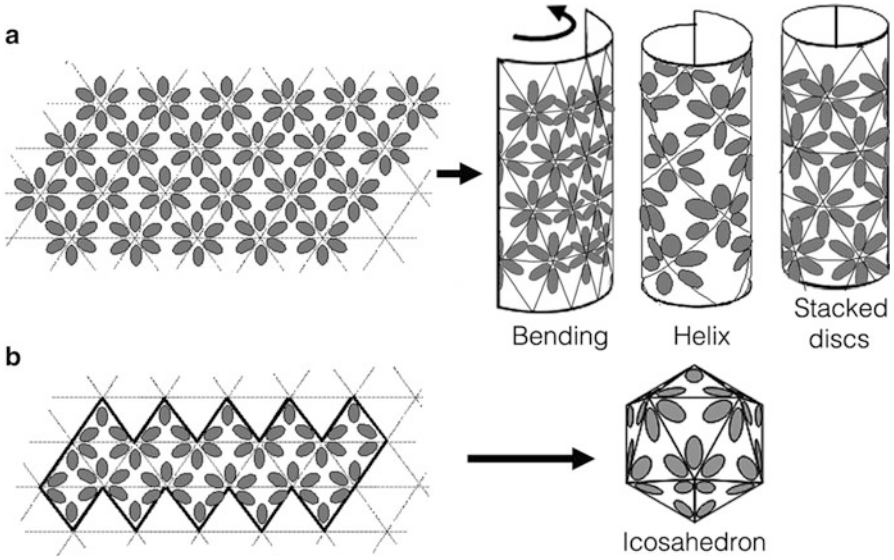


Fig. 2.3 Generation of curved structures such as a helix and an icosahedron from a planar hexagonal lattice. (**a**) Planar hexagonal lattice with subunits arranged in the triangular faces defined by the lattice vectors. (**b**) Icosahedron generated from a defined subset of faces derived from the hexagonal array

curvature in the lattice, which is closed as a polyhedron with 12 pentagonal vertices and 60 identical subunits. The icosahedron has a series of characteristic views depending on its orientation relative to the rotational symmetry of the solid, known as 532 symmetry. An icosahedron has six fivefold axes through the 12 vertices, ten threefold axes through the 20 triangular faces, and 15 twofold axes through the center of each edge (Fig. 2.4). The capsid described (with three identical subunits to form each triangular face) would be the simplest icosahedral capsid formed by 60 equivalent subunits.

2.4 Quasi-Equivalence Theory and Icosahedral Capsid Architecture

When the first icosahedral virus structures were characterized in some detail, it became evident that the limitation to 60 identical CP subunits was not compatible with the high molar mass and size of most viral capsids. The problem to be solved was how an icosahedral capsid could be made using more than 60 identical CP subunits. Based on the conceptual model of the hexagonal planar lattice, some hexamers must be added between pentamers to build larger icosahedra. In 1962, Caspar and Klug developed the quasi-equivalence theory, which constitutes the basis of structural studies of viruses [25].

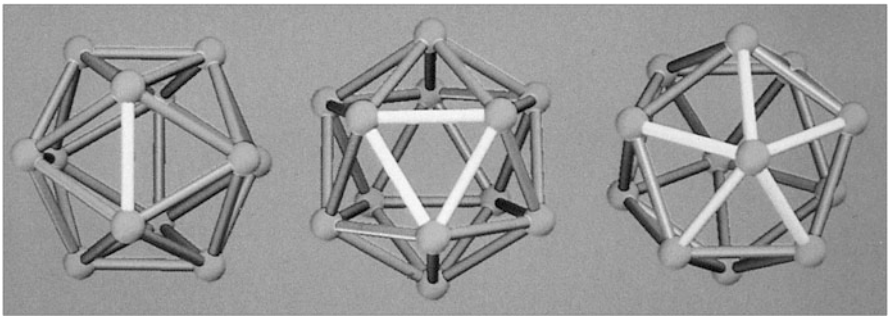


Fig. 2.4 Rotational icosahedral symmetry axes. The icosahedron viewed along two-, three- and fivefold symmetry axes (*left to right*). The twofold axis passes through the middle of each edge, the threefold axis passes through the center of each icosahedral face (an equilateral triangle), and the fivefold axis through the vertices of the icosahedron

The quasi-equivalence theory is based on the assumption that identical CP subunits interact forming nearly similar (quasi-equivalent) bonds for the construction of stable icosahedral capsids with multiples of 60 subunits. There is a degree of flexibility in the intersubunit contacts (conformational polymorphism), which involves minor variations in the 3D structure of the CP subunits. Each conformational state is termed a conformer. For example, the structure of the CP in pentamer subunits is not identical to that of the CP in the hexamer subunits. A number of different-sized icosahedra can be generated, because each has a unique ratio of hexamers to the 12 pentamers in the basic icosahedron. As indicated above, five or six subunits that interact closely are called pentameric or hexameric capsomers, respectively, or simply pentons and hexons. Given that the icosahedron is derived from the planar hexagonal lattice, not every combination of pentamers and hexamers is allowed; conversion of 12 non-adjacent hexamers into pentamers to form a closed shell must take place at regular, correctly separated intervals (Fig. 2.5a). The ratio that systematically defines all possible icosahedra from the hexagonal lattice is termed triangulation number (T), which matches the unit triangles of the elemental planar lattice contained in each triangular face of the icosahedron (facet) (Fig. 2.5a). The smallest viruses follow $T = 1$ geometry. The rules of quasi-equivalence allow only certain values of T; T is given by the formula $T = h^2 + k^2 + hk$, in which h and k are positive integers that define the pentamer positions in the original hexagonal lattice (Fig. 2.5a). Icosahedral viruses larger than $T = 1$ have T values of 3, 4, 7, 9, 12, 13, 16, 19, 21, 25, 27, 28, and so on.

The quasi-equivalence theory, combined with the concept of T, provides a systematic nomenclature with which to describe icosahedral capsid structure. Icosahedra fulfill Euler's polyhedral formula, faces (20) + vertices (12) = edges (30) + 2. Taking T into account, this formula becomes $20\ T + V = 30\ T + 2$. As there must always be 12 pentameric vertices, the total number of hexamers is 10 (T-1), and the total number of subunits is 60 T. T also describes the classes of conformers, or different quasi-equivalent (similar) environments occupied by a subunit (Fig. 2.5b); however, T does not describe the basic morphological unit or

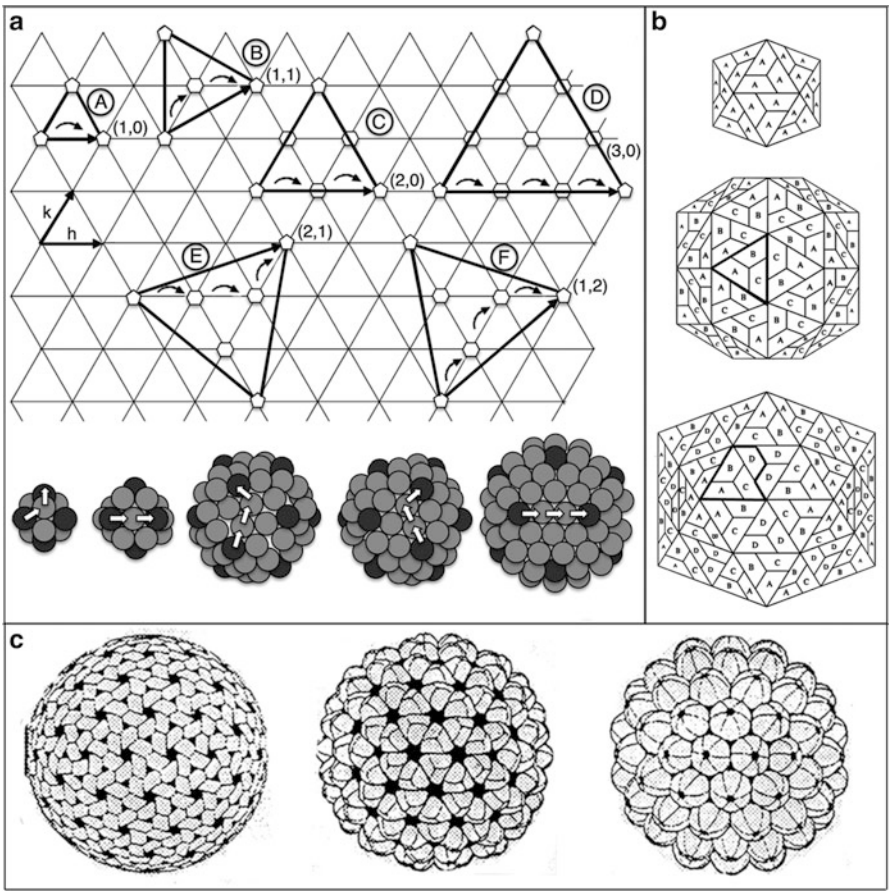


Fig. 2.5 The triangulation number T, and how it is determined in an icosahedral lattice. (**a**) The coordinate system of a planar hexagonal net with axes h and k crossing at a 60 º angle is formed by equilateral triangles; each triangle is formed by three protein subunits (as in Fig. 2.3), and specific hexamer-to-pentamer conversion introduces curvature into the hexamer sheet (see main text). Capsomers are located at the lattice intersection points; hexagons represent hexamers (or planar hexameric positions) and pentagons correspond to (*convex*) pentamers. Several triangular facets of an icosahedron are indicated (*bold outline*), and a facet edge of the icosahedron (*bold vector*) connects the two closest neighboring pentamers. The two unit vectors of the underlying hexagonal net, h and k, define the coordinates of the vector relative to its origin (positions at which hexamers are replaced by pentamers), and determine the parameters of the icosahedral capsid. The number of steps or spacings (*curved arrows*) from one structural unit (a hexamer or pentamer) to the adjacent one is therefore an integer. In example A, a pentamer is separated from the adjacent pentamer by one step along the h axis (*i.e.*, pentamers are in direct contact); hence, $h = 1$, $k = 0$, which corresponds to a $T = 1$ icosahedron (in which case 12 adjacent hexamers are converted to pentamers). Larger icosahedra are built by converting 12 nonadjacent hexamers to pentamers at regular, precisely-spaced intervals. In example B, one pentamer is separated from another by one step along the h axis and one step along the k axis; $h = 1$, $k = 1$, and $T = 3$. Note that T indicates the number of unit triangles contained in the icosahedral facet. In examples C and D, pentamers are separated by two and three steps along the direction of the h axis, and are $T = 4$ and $T = 9$,

repeated oligomer (other than hexamer/pentamer) that is actually used to build the capsid (the capsid building block), which must be determined empirically for each icosahedral virus (Fig. 2.5c).

Viral capsids with T numbers corresponding to $h \neq k$ (and both different from zero) result in skewed lattices with left- ($h > k$) or right-handed ($h < k$) possible configurations (see Fig. 2.5a). For example, for the skewed lattice with the smallest T number the left-handed (T = 7 *laevo* or 7 *l*) and right-handed (T = 7 *dextro* or 7*d*) configurations correspond to (h, k) arrangements (2, 1) and (1, 2), respectively. Besides the possible lattice handedness viral capsids are, as any protein assembly, enantiomeric structures. Therefore, the organization of capsomers (dimers or pentamers) or subunits might show a conspicuous handedness even in nonskewed capsids such as T = 1 or T = 3 (Fig. 2.5a).

Paradoxically, a T = 1 capsid with only one protein subunit is the least frequently observed quaternary structure in native viruses [13]. Only a few small ssRNA viruses with protein subunits of 20–40 kDa, such as satellite tobacco necrosis virus (STNV; 159-residue CP), are known to have this capsid type. Other T = 1 capsids are formed by ssDNA viruses such as parvoviruses and microviruses (*e.g.*, bacteriophage ϕX174) with CP as large as 75 kDa ([26] and references therein).

High-resolution structural studies have provided some clues as to how identical subunits adopt different conformations in a T > 1 icosahedral capsid [8, 9, 13, 27]. The distinct conformational states may be controlled by order/disorder of flexible regions of the protein (*e.g.*, in loops or N- and C- termini), dsRNA or ssRNA segments of the viral nucleic acid, metal ions, protons (pH), or combinations of any of

Fig. 2.5 (continued) respectively. A T = 4 lattice consists of 80 unit triangles, 240 subunits, and 4 quasi-equivalent protein subunits in the icosahedral asymmetric unit. Triangulated icosahedra might have local (or quasi) six-, three- and/or twofold symmetry. When the coordinates h or k are zero, or h = k, the icosahedral face is symmetric relative to the coordinate system (examples A, B, C and D). When h differs from k, the capsid is asymmetric or skewed, as in examples E and F. In a T = 7 icosahedron, the icosahedral face is turned relative to the hexagonal lattice, but the surfaces of the unit triangles gained or lost are compensated and T is a integer. The handedness of a T = 7 capsid is defined by the separation steps between the two closest pentamers, giving the maximum number of steps in the selected direction (this movement is similar to that of a knight in chess). In the T = 7laevo (left-handed) or 7 *l* (as in E), h = 2 and k = 1 (this step is leftward, along the direction of the k axis); similarly, in the T = 7dextro (right-handed) or 7*d* (as in F), k = 2 and h = 1 (a rightward step, along the direction of the h axis). The two T = 7 lattices are enantiomorphs, *i.e.*, mirror images of each other. The three-dimensional models (*bottom*) show how T number is determined; capsomers are shown as balls (*black*, *pentamer*; *grey*, *hexamer*). Note that capsid size is increased at higher T numbers, as capsomers are the same size. (**b**) Icosahedra corresponding to the triangulation numbers T = 1 (*top*), 3 (*middle*) and 4 (*bottom*); CP subunits are chemically identical (shown as trapezia). In a T = 1 capsid, every CP lies in an identical environment (labeled A); in a T = 3 capsid, the icosahedral asymmetric unit contains three CP subunits that occupy slightly different geometrical environments (A, B and C; one such unit is shown in thick lines); in a T = 4 capsid, each asymmetric unit contains four CP subunits (A, B, C and D). (**c**) Three different capsids with the same T number (T = 9) but different capsomers, viewed from a twofold axis; structural subunits are assembled without forming oligomers (*left*), as trimers (*middle*), or as pentamers and hexamers (*right*)

these [28]. The structural elements that participate in different, quasi-equivalent contacts at geometrically different positions in a $T > 1$ capsid are referred to as molecular switches. Removal of the molecular switching mechanism results in assembly of aberrant or non-native structures. In theory, quasi-equivalence involves small differences in subunit interactions and conformations; in practice, for capsids analyzed to date, viruses show broad variation, and the literature describes equivalent, quasi- and non-equivalent capsids. Whereas the structure of cowpea chlorotic mottle virus (CCMV) at atomic resolution showed that building blocks have quasi-equivalent contacts, with allowed variations in bond angles for noncovalent intersubunit interactions that stabilize oligomers, the two classes of dimeric building blocks in TBSV have completely different (strictly non-equivalent) bonding contacts [26].

Quasi-equivalence, which involves distinct spatial conformations, is supported by the intrinsic structural polymorphism of the CP. Structural polymorphism is also fundamental to the maturation process, since it allows in situ transient conformational changes throughout morphogenesis [29] (see Chap. 13). This strategy is optimal, as no additional genetic information is needed to generate these structural variants of the capsid.

Despite the flexibility and multiple possibilities that quasi-equivalence offers, capsid assembly frequently requires additional structural proteins. In many icosahedral viruses made of more than 60 subunits, quasi-equivalence in the strict sense of the term and molecular switches are not used at all. In these viruses, quasi-equivalent positions are occupied by different proteins, as in picornaviruses (see Chap. 10). The picornavirus capsid has 60 copies each of three subunits (VP0 (later processed to VP4 + VP2), VP1, VP3) that are synthesized as a precursor polyprotein (see Chap. 10). There is very little sequence similarity among them, but they all have the same fold, the canonical β-barrel. The additional genetic cost provides advantages difficult (if not impossible) to achieve only by conformational changes in a single protein, such as rapid acquisition of mutations to evade the immune system without altering capsid assembly [11–13]. When capsids are assembled from different gene products, T numbers are no longer valid, and pseudoT (P) numbers are used instead (for example, $P = 3$ in picornaviruses). Another well-known group with $P = 3$ capsids are the comoviruses, whose capsid has two subunits, a small subunit with a β-barrel domain and a large subunit formed by fusion of two β-barrels [18].

Molecular switches may be also insufficient to control assembly of capsids with large T numbers and/or when complex viruses are considered. These capsids may require one or more auxiliary proteins (scaffold, minor capsid or enzymatic proteins) that act as morphogenic factors to trigger structural changes [2, 30] (see Chap. 11).

There are many other variations on the icosahedral theme. For example, geminivirus particles consist of two fused $T = 1$ icosahedra, in which each lacks a pentameric cap [31]. In TBSV capsids, among others, particle curvature results from both pentamers and hexamers [10]. These variations are not restricted to capsids based on the tangential wedge-like β-barrel domain; the CP of HBV forms cores of varying T number and does not share this fold. The proportion of HBV $T = 3$ and $T = 4$ particles is affected by modifying the length of the C-terminus, which is located on the inner surface [32].

There are notable exceptions [30] that do not adhere to quasi-equivalence rules, such as the all-pentamer T = 7d capsid of papovaviruses (papilloma- and polyomaviruses), formed by 72 pentamers (360 subunits and six conformers, which would correspond to a disallowed T = 6) rather than 12 pentamers and 60 hexamers (420 subunits and seven conformers as predicted by quasi-equivalence) [33, 34]. In adenoviruses, which have a T = 25 shell (in fact, a pseudoT = 25, as pentameric subunits differ from hexameric subunits), each hexon is a trimer [14], but each subunit is formed by two similar domains with the β-barrel fold (see Chap. 11). A T = 25 capsid would have 25 conformers, but as the hexameric trimer is the basic unit, only four classes of (trimeric) conformers are actually needed. Another striking exception is the inner core of dsRNA viruses (formally, T = 1) formed by 12 decamers, in which dimers are the asymmetric unit; these capsids would correspond to the forbidden "T = 2" [35–37]. A number of studies have firmly established that this "T = 2" architecture is a common feature for dsRNA viruses ([38] and references therein).

Structural comparison of CP is used to establish evolutionary relatedness between viruses (or at least between their capsid genes) when sequence conservation is limited [13, 39, 40]; such comparisons led to detection of relationships among icosahedral viruses that infect hosts from three different domains of life (bacteria, eukarya and archaea). Icosahedral viruses are currently grouped in four separate lineages, each with common structural elements and assembly principles: these lineages are represented by the PRD1 bacteriophage, HK97 bacteriophage, bluetongue virus (BTV), and picornavirus-like viruses [41] (see also Chap. 7). The PRD1-like lineage includes adenovirus and viruses that infect archaea, such as STIV; the HK97-like lineage has the tailed dsDNA bacteriophages and herpesviruses, and the BTV-like lineage includes the icosahedral cores of dsRNA viruses and phage ϕ6. Although many viruses are not included in these four lineages, the number of folds that satisfy the assembly constraints for a viable viral shell is nonetheless thought to be limited.

2.5 Variations on the Icosahedral Capsid Theme: Multiple Layers and Prolate Icosahedra

Most animal viruses are icosahedral and follow the general principles described above, although many viruses have additional non-symmetric envelopes. The simplest icosahedral viruses include RNA and DNA viruses. Outstanding examples of RNA viruses belong to the *Reo-*, *Birna-*, *Calici-*, *Noda-*, *Tetra-* and *Picornaviridae* families; examples of DNA viruses belong to the *Parvo-*, *Papova-* and *Adenoviridae* families. Some icosahedral viruses have locally broken the symmetry at one of 12 fivefold vertices, due to the presence of the portal and tail complexes in place of a pentamer. To illustrate the variation in size and structural features among different viral capsids, we show 3D reconstructions of several distinct viruses calculated from cryo-EM images, from the 28 nm-diameter capsid of picornaviruses to the ~100 nm rotavirus capsid (Fig. 2.6). 3D maps are

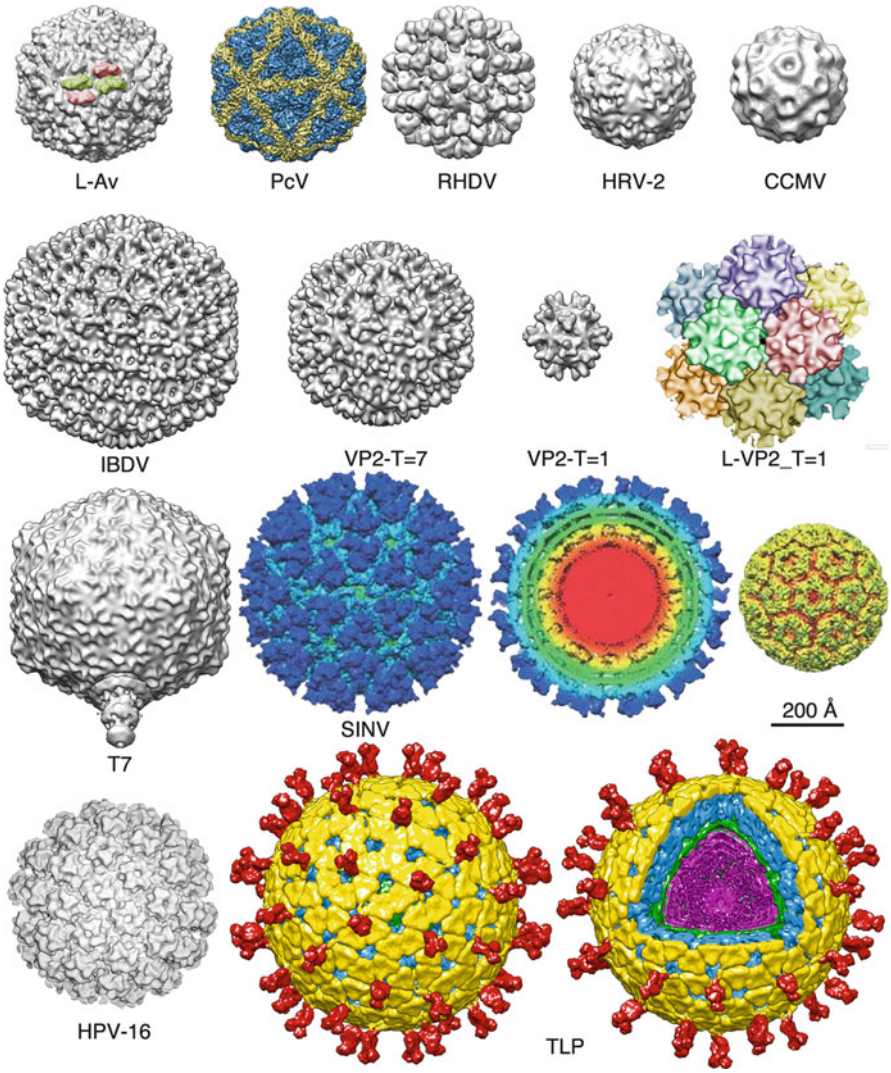


Fig. 2.6 Three-dimensional maps of icosahedral viruses calculated from two-dimensional cryo-electron microscopy images. L-Av, L-A virus of the yeast *Saccharomyces cerevisiae* with a T = 1 (or "T = 2") capsid (430 Å diameter); the basic unit is an asymmetric dimer (*green and red*) that assembles into decamers ([58], PDB entry 1M1C); PcV, Penicillium chrysogenum virus, T = 1, 400 Å; the basic unit is a monomer that has two similar domains (*yellow and blue*) (Adapted from [38]. Reproduced with permission); RHDV, rabbit hemorrhagic disease virus, T = 3, 400 Å; the basic unit is a dimer of the capsid protein (Adapted from [59]. Reproduced with permission); HRV-2 human rhinovirus 2, P = 3, 304 Å; the asymmetric unit is made of three different proteins (unpublished data; D. Blass and J.R. Castón); CCMV cowpea chlorotic mottle virus, T = 3, 300 Å (unpublished data; A. de la Escosura and J.R. Castón). CCMV capsid has 90 CP dimers (*protomers*) that assemble into 12 pentamers and 20 hexamers (*i.e.*, 32 capsomers or morphological units); IBDV infectious bursal disease virus, T = 13 *l* lattice, ~700 Å; the building block is a VP2 trimer (Adapted from [45, 60]. Reproduced with permission); VP2 can assemble into

intentionally presented at low to medium resolution (some are available at much higher resolution) to emphasize the capsomers (di-, tri-, penta- or hexamer) as well as the T numbers. Detection of specific capsomers implies stronger intra- than intercapsomeric interactions; isolated capsomers can be purified for structural and functional studies.

Many dsRNA viruses, including important human, animal and plant pathogens (rota-, reo-, and orbiviruses of the *Reoviridae* family) consist of a multilayered concentric capsid with icosahedral symmetry [42]. This complex organization is based on the role of each protein shell. Outer layers, usually a T = 13 shell, provide protection against chemical, physical and enzymatic damage and participate in receptor interaction and cell entry. The innermost core, a T = 1 shell with a protein dimer as the asymmetric unit, is essential for genome and replicative complex organization, participates in genome transcription and replication, and isolates the virus genome from host defense mechanisms. Fungal dsRNA viruses, which lack an extracellular route, have only a T = 1 capsid.

Capsids with larger T numbers might be considered advantageous, as formation of larger containers would in turn allow the incorporation of larger genomes; the concomitant packaging of specific viral components, such as polymerases and other enzymes, would reduce virus dependence on host cell machinery. Such a container, made at low economic cost, nonetheless presents enormous challenges. For example, in a quasi-equivalent T = 25 capsid that roughly emulates the adenovirus capsid, 1,500 copies of an individual protein must be laid accurately into the shell. In a relatively simpler T = 13 lattice, as in the birnavirus capsid, "only" 780 VP2 proteins must be arranged on the outer capsid surface (see Chap. 13). The VP2 subunit has three domains, an outer domain with an β-barrel fold in which β-strands are oriented radially, a shell domain with a β-barrel in tangential orientation, and an interior helix bundle contributed by the N- and C- termini [43]. The protein distribution problem is reduced in several ways: VP2 subunits form trimers, and five classes of these trimers are needed (rather than 13 conformers) to assemble into pentamers and hexamers. VP2 structural polymorphism is controlled by an inherent molecular switch, a C-terminal segment, that is absent in the mature VP2 form. Proteolytic processing of this switch is a finely tuned process in which viral

Fig. 2.6 (continued) T = 7 (VP2-T = 7; the left-handed form is chosen arbitrarily) and T = 1 (VP2-T = 1) capsids; T = 1 capsids self-assemble into a dodecahedron with 12 T = 1 capsids (L-VP2_T = 1); T7, T7 bacteriophage head, T = 7, ~600 Å [61]; the map shows a tail, located at the same vertex where the connector core is assembled; SINV Sindbis virus, an alphavirus with a double-layered particle, ~700 Å: outer T = 4 capsid (*blue*), a lipid bilayer (*green*), and an inner T = 4 capsid (*yellow*) (Adapted from [62]. With permission); HPV-16 human papillomavirus 16, T = 7d, ~600 Å; this capsid is made of 72 pentameric capsomers ([63]. Courtesy of BL Trus). TLP a rotavirus with a triple-layered particle, ~700 Å (without spikes): VP4 spikes (*red*), VP7 outermost T = 13 *l* layer (*yellow*), VP6 intermediate T = 13 *l* layer (*blue*), VP2 innermost T = 1 layer (*green*), and the internal density of dsRNA and polymerase complexes (*violet*) (unpublished data, courtesy of D. Luque and J.M. Rodríguez)

and cellular proteases participate; these proteolytic events take place in a procapsid-like structure, in which multiple scaffolding protein copies are essential for stabilizing or promoting interactions between precursor subunit trimers within hexamers [44–46].

The largest icosahedral viruses analyzed by cryo-EM include the insect larvae iridovirus CIV (T = 147, 185 nm diameter [47]), the algal dsDNA viruses *Paramecium bursaria chlorella* virus 1 (PBCV-1, T = 169d, 190 nm [48]) and *Phaeocystis pouchetii* virus (PpV01, T = 219d, 220 nm [49]), and the amoebal dsDNA mimivirus (estimated T between 972 and 1,200, 750 nm, [50]); all have a lipid membrane beneath the capsid and use minor CP for stabilization and to determine their assembly. Mimivirus can be "infected" by its own virus, Sputnik (with a T = 27 icosahedral capsid), and its genome size is similar to that of intracellular parasitic bacteria (see Chaps. 11 and 14).

A different strategy to increase the internal volume of the capsid is followed by some icosahedral-like viruses, which elongate the capsid along a fivefold axis with additional subunits in the central band or cylinder. This prolate icosahedral capsid is typical of bacteriophages such as ϕ29 [51] and T4 [52] (with a portal –the DNA packaging machine– and a tail –the DNA injecting machine– in a pentameric vertex), but is also found in several fungus, plant and animal viruses.

Consider an icosahedron made of three parts, the top and bottom caps (with five equilateral triangular facets) and the central band or cylinder (with ten triangles) that connect the caps. These capsids are characterized by two numbers, the T number and the elongation number Q (also referred to as T_{end} and T_{mid}, respectively). As the T number specifies the unit triangles of the triangular facet of an icosahedron (subdivisions into smaller triangles), the Q number also indicates the number of unit triangles in the stretched triangles of the central band, but can be any positive integer (Fig. 2.7). For icosahedral viruses, the total number of subunits is defined by n = 60 T; for prolate capsids, the total number of protein subunits is given by n = 30(T + Q). An isometric, icosahedral particle has Q = T; a prolate particle has Q > T. In the case of phage T4, the capsid can be defined by T = 13 *l* and Q = 20 [52]; in phage ϕ29, T = 3 and Q = 5 [51].

Assembly of prolate capsids based on the elongation of an icosahedron along a threefold or twofold axis of symmetry has also been suggested for some bacilliform viruses such as AMV; geometrical rules have been established for their construction [53].

2.6 Helical Capsids

Helical symmetry facilitates assembly of nucleoprotein complexes without size limitations, as variable-sized genome could, in principle, be coated by identical protein subunits that surround it, forming a cylinder with helical symmetry ([54], and references therein). Although helical symmetry simplifies capsid assembly and offers flexibility for genome packaging, this assembly is relatively infrequent. Less

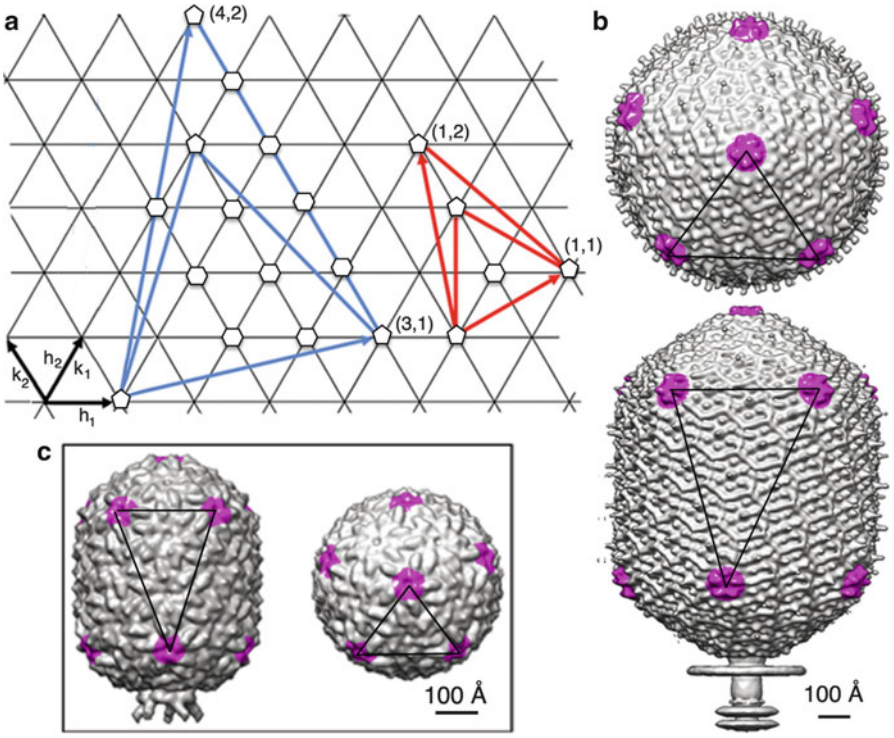


Fig. 2.7 Determination of the triangulation number (T) and the elongation number (Q) in a prolate icosahedral lattice. (**a**) A planar equilateral triangular lattice (as in Fig. 2.5) is used to show the triangulation numbers, T (or T_{end}) and Q (or T_{mid}), of prolate capsids; two unit vector systems, (h_1,k_1) and (h_2,k_2), are used to represent the equilateral triangular facets at the end caps and the elongated triangular facets at the central band (see text). The large equilateral triangle (*blue*) represents an end cap facet of a bacteriophage T4 head, and is defined by the (h_1, k_1) coordinate point [the (h_1, k_1) lattice point corresponds to a (h, k) point in Fig. 2.5]; $h_1 = 3$, $k_1 = 1$, and $T_{end} = h_1^2 + h_1k_1 + k_1^2 = 13$ *l*. The large triangle with unequal sides (*blue*) represents the phage T4 midsection facets and is described by two vectors, one from the origin to the lattice points (h_1, k_1) and the other to (h_2,k_2); $h_2 = 4$, $k_2 = 2$, and the triangulation of the extended facet is $Q = T_{mid} = h_1h_2 + h_1k_2 + k_1k_2 = 20$. For bacteriophage ϕ29, the equilateral facets at the end caps (*red*) are defined by $h_1 = 1$, $k_1 = 1$, and $T_{end} = 3$, and the stretched facets at the midsection triangles by $h_2 = 1$, $k_2 = 2$, and $Q = 5$. (**b**) Surface-shaded representation of the 3D cryo-EM map of the T4 bacteriophage head ([52], cryo-EM map deposited at the EMDataBank; ID EMD_1075), viewed along the fivefold axis of an end cap (*top*) and perpendicular to the fivefold axis (*bottom*). The equilateral and stretched triangles show the icosahedral equilateral and elongated facets defined by pentameric vertices (*pink*). **c** 3D map of ϕ29 bacteriophage prohead ([51], EMD_1117) showing equivalent views as for T4 phage

than 10 % of viral families have helical symmetry and most of them are plant viruses and bacteriophages. The most evident problems associated with this geometry are the disadvantageous surface/volume ratio and the structural instability inherent to long, thin structures. The paradigm of helical viruses is the tobacco mosaic virus (TMV), the first virus to be identified. A helix is defined by its pitch

(P), which includes two parameters: the number of subunits per helix turn (μ) and axial rise per subunit (ρ); thus, P = μ x ρ. For TMV, pitch is 23 Å, μ = 16.3 and ρ = 1.4 Å. TMV particles are rigid, right-handed, rod-like structures, 300 nm long and 18 nm in diameter, with a central 4 nm-diameter hole. Other helical viruses however, such as the Inoviridae (phages M13 and fd) are longer and flexible.

The genome of many animal negative-sense ssRNA viruses is also protected by a helical capsid which, together with the enclosed nucleic acid, forms the nucleocapsid, which is in turn invariably surrounded by a lipid envelope. Unlike many phage and plant viruses, no known animal viruses have a nonenveloped helical nucleocapsid. To provide some examples, rhabdoviruses (rabies and stomatitis vesicular viruses) and paramyxovirus (Sendai, mumps and measles viruses) have left-handed helical nucleocapsids surrounded by a lipid envelope. Orthomyxovirus (influenza virus) have multisegment helical nucleocapsids, referred to as ribonucleoproteins (RNP). As for most enveloped viruses, the influenza virus nucleocapsids are surrounded by an amorphous layer formed by the matrix protein, which also interacts with the external lipid envelope.

2.7 The Viral Nucleic Acid Inside

The mechanisms of nucleic acid packaging and related aspects will be reviewed in depth in Chap. 12. We will focus here only on some general aspects of the organization of the viral nucleic acid in the capsid from the point of view of virus architecture. Whereas the assembly of CP subunits to build a viral capsid follows a few well-defined strategies, and in most cases yields a unique geometric solution as outlined above, the organization of the nucleic acid inside the capsid is more diverse, and is greatly conditioned by capsid geometry. In viruses with helical symmetry, such as the rigid rod-shaped plant viruses (*e.g.*, TMV), ssRNA is co-assembled with the CP, integrated such that it strictly follows the geometrical characteristics imposed by the CP assembly. Other DNA-containing helical viruses (like phage M13) also show a close structural relationship between nucleic acid and protein. In this case, the fit is less tight than in TMV-like viruses, and is based on defined contacts between the DNA and the inner walls of the helical protein shell. A different type of viral helical protein-nucleic acid arrangement is found in flexible nucleoprotein complexes characteristic of Mononegavirales (*Borna-*, *Filo-*, *Paramyxo-* and *Rhabdoviridae* families). These helical assemblies, which are enclosed within lipid-based envelopes, work as an active complex for replication and transcription.

Nucleic acid arrangement in icosahedral capsids shows a wide variety of solutions, from partly disordered to highly compact, quasi-crystalline organization. In many ssRNA viruses, most of the RNA (up to 80 %) is disordered, as the molecules interact with the icosahedral capsid and use only small nucleotide stretches with specific capsid symmetries (as in the case of satellite tobacco mosaic virus or flock house virus. These ordered nucleotide stretches are important, as they

are probably involved in the assembly of the CP subunits by directing and/or stabilizing the assembly intermediates (Chap. 12). At the other extreme, dsDNA bacteriophages show very high nucleic acid density inside the capsid (>500 mg/ml). In these cases, viruses have incorporated a sophisticated machinery to package the DNA molecule as a tight coil that forms concentric shells inside the capsid, with the DNA strands arranged in a parallel pseudo-hexagonal crystalline array (Chap. 12).

2.8 Basic Architecture of Enveloped Viruses

We will refer here briefly to the general structural organization of enveloped viruses in which a lipid membrane constitutes the outermost structural layer. For detailed descriptions of different aspects of the structure and function of enveloped virus, see Chaps. 11, 15, and 16.

Capsid or nucleocapsid assembly is not necessarily the final step in virion morphogenesis, as some viral capsids or nucleocapsids become surrounded by one or several membrane envelopes of cellular origin, in which viral proteins are embedded. The lipid bilayer is acquired during the budding from a cell membrane, which can be the plasma membrane during escape from host cells, or membrane-delimited structures in eukaryotic cells such as the nuclear membrane, endoplasmic reticulum, Golgi apparatus and vacuoles. Enveloped viruses can exit the cell without causing cell death, since the budding process does not alter cell membrane integrity, and can cause either persistent or latent infections. Viral glycoproteins are inserted in the membranes *via* the compartmentalization routes and constant flux of vesicular trafficking. Glycoproteins have an external domain (ectodomain) with sugar chains, which is anchored in the membrane by a transmembrane hydrophobic segment (typically α-helix) that terminates in a small domain (endodomain) facing the cytoplasm. Viral glycoprotein monomers associate to form multimeric spikes (visible by electron microscopy), which are the functional state for at least four activities: identification and binding to cell receptors, membrane fusion, transcription activation, and receptor destruction. These functions can be carried out by one or several glycoproteins. Envelopes studded with glycoproteins usually have minor proteins that make ionic channels. Finally, matrix proteins are internal proteins that mediate association of glycoprotein endodomains with the nucleocapsid (see Chap. 16).

Membrane acquisition differs among viruses. Alpha- and flaviviruses, which have a positive-sense ssRNA genome, have a nucleocapsid that is assembled independently in the host cytoplasm and subsequently interacts with the glycoprotein endodomains. In alphaviruses (of the *Togaviridae* family), lateral interactions among glycoproteins exclude cell membrane proteins, resulting in a three-layered capsid: a glycoprotein outer icosahedral layer, a lipid bilayer, and a protein inner icosahedral layer (both icosahedral shells are in register with a $T = 4$ geometry [55]). Flaviviruses have an outer ordered protein layer above a polygonal lipid bilayer, and an inner disordered nucleocapsid; the outer capsid is formed by 90 closely packed glycoprotein dimers that lie flat on the particle surface, rather than forming protruding

spikes. Dimer arrangement differs from a true T = 3 shell, however, and the capsid should be considered a T = 1 capsid with three monomers per asymmetric unit ([56] and references therein). Whereas budding takes place in the plasma membrane for alphaviruses, flavivirus buds in the endoplasmic reticulum lumen.

Contrary to the well ordered, quasi-crystalline arrangement found in icosahedral capsids, most enveloped viruses lack regular organization in the membrane. Orthomyxoviruses (influenza virus) have an additional protein, the matrix protein, which mediates interactions between helical nucleocapsids (in eight segments) and the viral envelope. Retroviruses have a similar membrane-associated matrix protein. Influenza virus membrane has two glycoproteins, neuraminidase, which cleaves sialic acid groups from receptor glycoproteins, and hemagglutinin, a trimer that contains the binding site for the virus receptor and is responsible for viral and cell membrane fusion. Rhabdo- and paramyxoviruses are similar to orthomyxoviruses, except that the genome is a single helical nucleocapsid. Bunyaviruses have a tripartite genome and are very flexible, since they lack matrix proteins; they have glycoprotein spikes, some ordered in a T = 12 capsid, some locally ordered, and some lacking any order. Herpesvirus virions comprise a T = 16 icosahedral capsid (125 nm diameter) and interact with the envelope (derived from the trans-Golgi network) through a complex asymmetric layer with many viral (and a few cellular) proteins, called the tegument (virion size, 200 nm diameter). The herpesvirus envelope has 12 different glycoproteins and about 800 spikes per virion. Poxviruses (or brick-shaped viruses) such as vaccinia virus encompass multiple structural elements, and are described in detail in Chap. 11.

Retroviruses, such as the lentivirus HIV, are enveloped pleomorphic virions. HIV is about 100 nm in diameter and has two copies of a positive sense ssRNA genome associated with the nucleoprotein basic protein (NC). The dimeric ribonucleoprotein complex is inside a protein shell built by about 1,500 copies of a single protein, CA, which is assembled into a cone-shaped capsid in the mature HIV virion. There is a matrix protein (MA) layer that interacts with the inner leaflet of the viral envelope (MA is a peripheral membrane protein). MA, CA and NC are synthesized as a single precursor, Gag, that is arranged radially in immature, incomplete spherical particles. Immature HIV virions bud at lipid rafts; these membranes have viral spikes formed by the surface (gp120) and transmembrane (gp41) glycoproteins, which are also initially fused. Virions mature after budding by the concomitant activation of the viral protease. CA forms hexamers on a variably curved hexagonal lattice closed by insertion of 12 pentamers, seven at the wide and five at the narrow end of the fullerene cone. CA has two domains, CA^{NTD} and CA^{CTD}; in the mature capsid, CA^{NTD} forms hexameric and pentameric rings, whereas CA^{CTD} dimeric contacts, at base of the protein shell, connect neighboring rings ([57] and references therein) (see also Chap. 5).

Acknowledgements We thank Daniel Luque for preparing several figures used in this chapter and Catherine Mark for editorial help. This work was supported by grants BFU2011-25902 (to JRC) and BFU2011-29038 (to JLC) from the Spanish Ministry of Science and Innovation.

References and Further Reading

1. Harrison SC (2007) Principles of virus structure. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Strauss SE (eds) Fields virology, vol 1. Lippincott Williams & Wilkins, Philadelphia, pp 59–98
2. Dokland T (2000) Freedom and restraint: themes in virus capsid assembly. Structure Fold Des 8:R157–R162
3. Flint SJ, Enquist LW, Krug RM, Racaniello VR, Skalka AM (2000) Principles of virology. Molecular biology, pathogenesis, and control. ASM Press, Washington, DC
4. Cann AJ (2012) Principles of molecular virology. Elsevier, Amsterdam
5. Baker TS, Olson NH, Fuller SD (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol Mol Biol Rev 63:862–922
6. Prasad BV, Schmid MF (2012) Principles of virus structural organization. Adv Exp Med Biol 726:17–47
7. Hendrix RW, Johnson JE (2012) Bacteriophage HK97 capsid assembly and maturation. Adv Exp Med Biol 726:351–363
8. Chang J, Liu X, Rochat RH, Baker ML, Chiu W (2012) Reconstructing virus structures from nanometer to near-atomic resolutions with cryo-electron microscopy and tomography. Adv Exp Med Biol 726:49–90
9. Zhou ZH (2011) Atomic resolution cryo electron microscopy of macromolecular complexes. Adv Protein Chem Struct Biol 82:1–35
10. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G (1978) Tomato bushy stunt virus at 2.9 A resolution. Nature 276:368–373
11. Hogle JM, Chow M, Filman DJ (1985) Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358–1365
12. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ, Johnson JE, Kamer G, Luo M, Mosser AG et al (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317:145–153
13. Rossmann M, Johnson J (1989) Icosahedral RNA virus structure. Annu Rev Biochem 58:533–573
14. Roberts MM, White JL, Grutter MG, Burnett RM (1986) Three-dimensional structure of the adenovirus major coat protein hexon. Science 232:1148–1151
15. Nandhagopal N, Simpson AA, Gurnon JR, Yan X, Baker TS, Graves MV, Van Etten JL, Rossmann MG (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc Natl Acad Sci U S A 99:14758–14763
16. Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martin C, Burnett RM, Stuart DI, Bamford DH, Bamford JK (2004) Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432:68–74
17. Fu CY, Johnson JE (2012) Structure and cell biology of archaeal virus STIV. Curr Opin Virol 2:122–127
18. Lomonossoff GP, Johnson JE (1991) The synthesis and structure of comovirus capsids. Prog Biophys Mol Biol 55:107–137
19. Bahar MW, Graham SC, Stuart DI, Grimes JM (2011) Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. Structure 19:1011–1020
20. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289:2129–2133
21. Veesler D, Johnson JE (2012) Virus maturation. Annu Rev Biophys 41:473–496
22. Zlotnick A (2005) Theoretical aspects of virus capsid assembly. J Mol Recognit 18:479–490
23. Johnson JE (2008) Multi-disciplinary studies of viruses: the role of structure in shaping the questions and answers. J Struct Biol 163:246–253
24. Crick FH, Watson JD (1956) Structure of small viruses. Nature 177:473–475
25. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harbor Symp Quant Biol 27:1–24

26. Liljas L (1986) The structure of spherical viruses. Prog Biophys Mol Biol 48:1–36
27. Johnson JE, Speir JA (1997) Quasi-equivalent viruses: a paradigm for protein assemblies. J Mol Biol 269:665–675
28. Johnson JE (1996) Functional implications of protein-protein interactions in icosahedral viruses. Proc Natl Acad Sci USA 93:27–33
29. Chen DH, Baker ML, Hryc CF, DiMaio F, Jakana J, Wu W, Dougherty M, Haase-Pettingell C, Schmid MF, Jiang W, Baker D, King JA, Chiu W (2011) Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. Proc Natl Acad Sci USA 108:1355–1360
30. Steven A, Trus B, Booy F, Cheng N, Zlotnick A, Castón J, Conway J (1997) The making and breaking of symmetry in virus capsid assembly: glimpses of capsid biology from cryoelectron microscopy. FASEB J 11:733–742
31. Zhang W, Olson NH, Baker TS, Faulkner L, Agbandje-McKenna M, Boulton MI, Davies JW, McKenna R (2001) Structure of the Maize streak virus geminate particle. Virology 279:471–477
32. Zlotnick A, Cheng N, Conway JF, Booy FP, Steven AC, Stahl SJ, Wingfield PT (1996) Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Biochemistry 35:7412–7421
33. Rayment I, Baker TS, Caspar DL, Murakami WT (1982) Polyoma virus capsid structure at 22.5 A resolution. Nature 295:110–115
34. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC (1991) Structure of simian virus 40 at 3.8-A resolution. Nature 354:278–284
35. Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Zientara S, Mertens PP, Stuart DI (1998) The atomic structure of the Bluetongue virus core. Nature 395:470–478
36. Reinisch KM, Nibert ML, Harrison SC (2000) Structure of the Reovirus core at 3.6 Å resolution. Nature 404:960–967
37. Naitow H, Tang J, Canady M, Wickner RB, Johnson JE (2002) L-A virus at 3.4 a resolution reveals particle architecture and mRNA decapping mechanism. Nat Struct Biol 9:725–728
38. Luque D, Gonzalez JM, Garriga D, Ghabrial SA, Havens WM, Trus B, Verdaguer N, Carrascosa JL, Caston JR (2010) The T = 1 capsid protein of penicillium chrysogenum virus is formed by a repeated helix-rich core indicative of gene duplication. J Virol 84:7256–7266
39. Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? Curr Opin Struct Biol 15:655–663
40. Baker ML, Jiang W, Rixon FJ, Chiu W (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. J Virol 79:14967–14970
41. Abrescia NG, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral universe. Annu Rev Biochem 81:795–822
42. Pesavento JB, Crawford SE, Estes MK, Prasad BV (2006) Rotavirus proteins: structure and assembly. Curr Top Microbiol Immunol 309:189–219
43. Coulibaly F, Chevalier C, Gutsche I, Pous J, Navaza J, Bressanelli S, Delmas B, Rey FA (2005) The birnavirus crystal structure reveals structural relationships among icosahedral viruses. Cell 120:761–772
44. Saugar I, Luque D, Ona A, Rodriguez JF, Carrascosa JL, Trus BL, Caston JR (2005) Structural polymorphism of the major capsid protein of a double-stranded RNA virus: an amphipathic alpha helix as a molecular switch. Structure 13:1007–1017
45. Luque D, Saugar I, Rodriguez JF, Verdaguer N, Garriga D, Martin CS, Velazquez-Muriel JA, Trus BL, Carrascosa JL, Caston JR (2007) Infectious bursal disease virus capsid assembly and maturation by structural rearrangements of a transient molecular switch. J Virol 81:6869–6878
46. Irigoyen N, Caston JR, Rodriguez JF (2012) Host proteolytic activity is necessary for infectious bursal disease virus capsid protein assembly. J Biol Chem 287:24473–24482
47. Yan X, Yu Z, Zhang P, Battisti AJ, Holdaway HA, Chipman PR, Bajaj C, Bergoin M, Rossmann MG, Baker TS (2009) The capsid proteins of a large, icosahedral dsDNA virus. J Mol Biol 385:1287–1299
48. Zhang X, Xiang Y, Dunigan DD, Klose T, Chipman PR, Van Etten JL, Rossmann MG (2011) Three-dimensional structure and function of the *Paramecium bursaria chlorella* virus capsid. Proc Natl Acad Sci U S A 108:14837–14842

49. Yan X, Chipman PR, Castberg T, Bratbak G, Baker TS (2005) The marine algal virus PpV01 has an icosahedral capsid with T = 219 quasisymmetry. J Virol 79:9236–9243
50. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG (2009) Structural studies of the giant mimivirus. PLoS Biol 7:e92
51. Tao Y, Olson NH, Xu W, Anderson DL, Rossmann MG, Baker TS (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell 95:431–437
52. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG (2004) Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci U S A 101:6003–6008
53. Luque A, Reguera D (2010) The structure of elongated viral capsids. Biophys J 98:2993–3003
54. Stubbs G, Kendall A (2012) Helical viruses. Adv Exp Med Biol 726:631–658
55. Zhang R, Hryc CF, Cong Y, Liu X, Jakana J, Gorchakov R, Baker ML, Weaver SC, Chiu W (2011) 4.4 A cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. EMBO J 30:3854–3863
56. Mukhopadhyay S, Kuhn RJ, Rossmann MG (2005) A structural perspective of the flavivirus life cycle. Nat Rev Microbiol 3:13–22
57. Ganser-Pornillos BK, Yeager M, Pornillos O (2012) Assembly and architecture of HIV. Adv Exp Med Biol 726:441–465
58. Castón JR, Trus BL, Booy FP, Wickner RB, Wall JS, Steven AC (1997) Structure of L-A virus: a specialized compartment for the transcription and replication of double-stranded RNA. J Cell Biol 138:975–985
59. Luque D, Gonzalez JM, Gomez-Blanco J, Marabini R, Chichon J, Mena I, Angulo I, Carrascosa JL, Verdaguer N, Trus BL, Barcena J, Caston JR (2012) Epitope insertion at the N-terminal molecular switch of the rabbit hemorrhagic disease virus T = 3 capsid protein leads to larger T = 4 capsids. J Virol 86:6470–6480
60. Castón JR, Martínez-Torrecuadrada JL, Maraver A, Lombardo E, Rodríguez JF, Casal JI, Carrascosa JL (2001) C terminus of infectious bursal disease virus major capsid protein VP2 is involved in definition of the T number for capsid assembly. J Virol 75:10815–10828
61. Agirrezabala X, Martin-Benito J, Caston JR, Miranda R, Valpuesta JM, Carrascosa JL (2005) Maturation of phage T7 involves structural modification of both shell and inner core components. EMBO J 24:3820–3829
62. Jose J, Snyder JE, Kuhn RJ (2009) A structural and functional perspective of alphavirus replication and assembly. Future Microbiol 4:837–856
63. Buck CB, Trus BL (2012) The papillomavirus virion: a machine built to hide molecular Achilles' heels. Adv Exp Med Biol 726:403–422

Further Reading

Chiu W, Burnnett RM, Garcea RL (eds) (1997) Structural virology of viruses. Oxford University Press, New York
Chiu W, Johnson JE (eds) (2003) Virus structure. Advances in protein chemistry, vol 64. Academic Press, Amsterdam
Mahy BWJ, van Regenmortel MHV (eds) (2008) Encyclopedia of virology. Elsevier, Amsterdam
Patton JT (ed) (2008) Segmented double-stranded RNA viruses, structure and molecular biology. Caister Academic Press, Norfolk
Rossmann MG, Rao VB (eds) (2012) Viral molecular machines. Adv Exp Med Biol, vol 726. Springer, New York

Also especially recommended for further reading are references [5, 6, 9, 21, 30, 41] *listed above*.

Part II
Determination of the Structure and Physical Properties of Viruses

Chapter 3
Conventional Electron Microscopy, Cryo-Electron Microscopy and Cryo-Electron Tomography of Viruses

José R. Castón

Abstract Electron microscopy (EM) techniques have been crucial for understanding the structure of biological specimens such as cells, tissues and macromolecular assemblies. Viruses and related viral assemblies are ideal targets for structural studies that help to define essential biological functions. Whereas conventional EM methods use chemical fixation, dehydration, and staining of the specimens, cryo-electron microscopy (cryo-EM) preserves the native hydrated state. Combined with image processing and three-dimensional reconstruction techniques, cryo-EM provides 3D maps of these macromolecular complexes from projection images, at subnanometer to near-atomic resolutions. Cryo-EM is also a major technique in structural biology for dynamic studies of functional complexes, which are often unstable, flexible, scarce or transient in their native environments. As a tool, cryo-EM complements high-resolution techniques such as X-ray diffraction and NMR spectroscopy; these synergistic hybrid approaches provide important new information. Three-dimensional cryo-electron tomography goes further, and allows the study of viruses not only in their physiological state, but also in their natural environment in the cell, thereby bridging structural studies at the molecular and cellular levels.

Keywords Capsid • Cryo-electron microscopy • Cryo-electron tomography • Electron microscopy • Fourier transform • Image processing • Resolution • Three-dimensional reconstruction • Viral macromolecular assembly • Virus

J.R. Castón (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jrcaston@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_3,

Abbreviations

2D	Two-dimensional, two dimensions
3D	Three-dimensional, three dimensions
3DR	Three-dimensional reconstruction
CCD	Charge-coupled device
cryo-EM	Cryo-electron microscopy
cryo-ET	Cryo-electron tomography
CTF	Contrast transfer function
EM	Electron microscopy, electron microscope
FEG	Field emission gun
FSC	Fourier shell correlation
FT	Fourier transform
SEM	Scanning electron microscopy
SSE	Secondary structure element
TEM	Transmission electron microscopy

3.1 Introduction

Microscopy is probably the technique with the greatest influence in the history of biology. Optical (or light) microscopy was introduced to biology in the seventeenth century by Antoni van Leeuwenhoek, who described bacteria and protozoa using a microscope able to enlarge image size approximately 300 times. Light microscopy uses visible light and a system of lenses to magnify images, providing intuitive information on morphology and general structure of the biological model under study. The end of the nineteenth and beginning of the twentieth centuries saw the discovery of electrons and X rays; both provided new alternatives for the observation of matter and stimulated the development of the electron microscope and the X-ray diffractometer. In 1931, Ernst Ruska and colleagues invented the electron microscope (Ruska obtained the Nobel Prize in Physics in 1986). One of the early micrographs taken in the 1930s was of bacterial viruses or bacteriophages, called "tiny hostile bacteria".

It took 300 years to improve the simple light microscope, but less than 40 years to perfect the electron microscope. With electron microscopy (EM), scientists were provided with a powerful method that extended the range of observable structures far beyond the limits imposed by the physics of visible light. Our understanding of cell structure and tissues is based in numerous EM contributions. As in many other fields in biology, viruses have had a decisive role in the development of EM, as they constitute simplified tools with which to standardize preparation methods in

biological applications. In turn, EM has made outstanding contributions to understand virus structure and its relationship with biological function.

Biological EM might be considered a descriptive technique in its origins. The questions addressed by EM have evolved together with our concept of the molecular reality of the cell [1]. The experience accumulated over many decades in the multidisciplinary field of biology has shown that each molecular event is more than the sum of its parts. This is reflected, for example, in the regulation and integration of physiological signals, currently termed "systems biology". Genomics has provided a complete list of the macromolecules in a cell, which led to structural genomics, which in turn provides the basic principles of the structure of proteins, of DNA and of RNA. The basic cellular entities that carry out fundamental biological processes are nonetheless multiprotein assemblies [2]. These macromolecular complexes (some of which may contain nucleic acids) respond to factors that determine their association, disassembly or signal transmission, leading finally to specific molecular events.

To decipher biological processes, the organization of macromolecules and components of macromolecular complexes must be established. X-ray crystallography (Chap. 4) and nuclear magnetic resonance (NMR) spectroscopy (Chap. 5) have provided the atomic structures of the modules or subunits of these molecular machines. Given the difficulties in producing sufficient amounts of these complexes and in their crystallization, the number of structures of large macromolecular complexes that have been determined at near-atomic resolution (by X-ray crystallography) is limited compared to the number of protein structures available. EM acts as a link between structural cell biology and high-resolution structural molecular biology. Over time, descriptive EM has given way to a quantitative approach [3]. Three-dimensional (3D) density maps can be synthesized from two-dimensional (2D) images, as electron micrographs are 2D projections of the 3D (multi-) macromolecular assembly under study [4]. EM images nonetheless have deficiencies, as the electron microscope is not a perfect instrument. These problems can be reduced by subsequent image processing of the micrographs. This progress is reflected in the relatively large number of biological structures that have been resolved thus far to resolutions in the 3.5–4.5-Å range.

EM facilitates the understanding of an atomic structure in its biological context (at lower resolution), particularly when macromolecular assemblies and the relationships of their components are involved. The frontiers between NMR spectroscopy, X-ray crystallography and EM are becoming diffuse, as they form a smooth continuum for visualizing macromolecular structures [5] (see also Chaps. 4, 5 and 7). The emerging relationship among these approaches is a complementary one that can overcome the limitations of each method alone; for instance, atomic models can be fitted accurately into EM-derived maps of larger assemblies [6] (see Chap. 7). Viruses are excellent models of nanomachines; their capsids and other viral protein assemblies have provided some of the most striking images using this hybrid approach [7]. In addition to structure and assembly, the form of EM termed cryo-electron microscopy (cryo-EM) also allows dynamic studies of complexes, which are often unstable, scarce or transient, in their native environment.

A structural description of a macromolecular assembly in its different functional states facilitates a mechanistic understanding of the corresponding processes [8].

In this chapter, we describe the principles of EM, with examples of its applications to the study of some viruses. Most sections in this chapter outline the basis of 3D cryo-EM (cryo-EM plus image processing), illustrating its potential in the marriage with X-ray crystallography (see also Chap. 7). The powerful technique of 3D cryo-electron tomography (cryo-ET), which allows the study of viruses in their natural environment in the cell, is also described. The examples mentioned may give the reader a perspective as to how this approach can be applied to the study of large viral complexes and allow determination of their overall shape, the location of subunits and domains within them, and identification of conformational changes that accompany intermediates in their functional cycle. Moreover, even today in the age of molecular diagnosis, EM remains a mainstay in detecting the causative organism in outbreaks of new and unusual diseases; EM does not require specific reagents to be able to recognize a pathogenic agent [9].

3.2 Transmission Electron Microscopy of Viruses

3.2.1 Basic Concepts and General Experimental Design

The physical basis of EM is the dual particle-wave nature of the electron and its charge. Electrons can be accelerated by electric fields reaching high velocities and accordingly much shorter associated wavelengths than visible light photons, leading to much improved resolution compared with an optical microscope. Electrons are provided by an emission source, such as a heated tungsten or lanthanum hexaboride filament or, alternatively, by a field emission gun (FEG). A FEG generates a much brighter electron beam with better spatial coherence than a thermal ionic electron source. Electrons are accelerated by high voltage, and focused by electromagnetic lenses (condenser lenses) to make a collimated (highly parallel) beam that interacts with the specimen. The interior of the microscope column must be under a high vacuum system to reduce the frequency of electron collision with gas atoms to negligible levels in the path of the electrons; this system also prevents degradation of the coherence and monochromaticity of the electron beam (Fig. 3.1). The sample is inserted in the column through an airlock.

There are two basic types of electron microscopes, the transmission electron microscope (TEM) and the scanning electron microscope (SEM). TEM projects electrons through a very thin specimen; the beam interacts with the specimen and the transmitted electrons are focused by electromagnetic lenses (objective lens and projector lenses) onto a detector to produce a magnified 2D image of the specimen [10]. The electron beam covers a wide area of the specimen. Detectors are a fluorescent screen, a photographic film or a charge-coupled device (CCD). In addition, direct electron detectors (without scintillators to convert the electrons

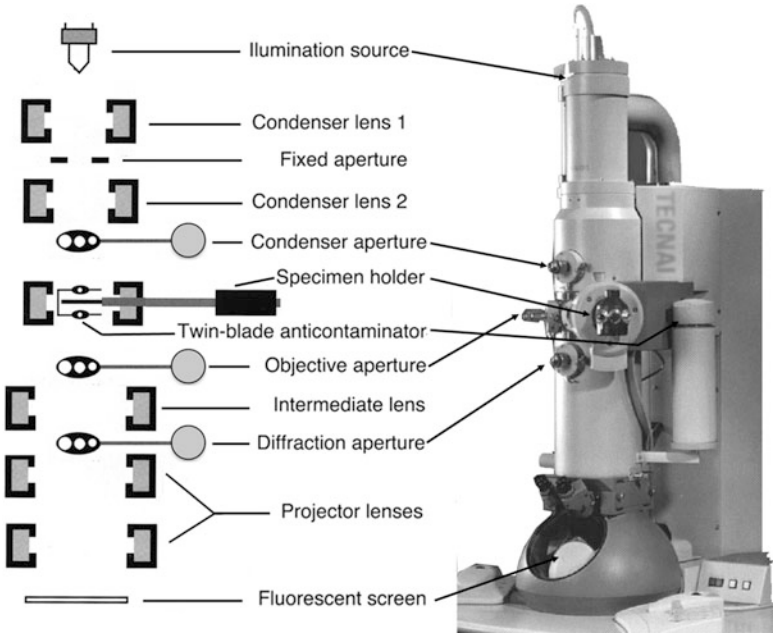


Fig. 3.1 Major components of a transmission electron microscope. Diagram of major components in the column of a modern transmission (cryo-)electron microscope

into photons) are being explored, as they would be superior recording devices for EM images. The brightness of a particular area of the image is proportional to the number of electrons that are transmitted through the specimen.

SEM produces an image that gives the impression of three dimensions. In this microscope, the electron beam is focused to a fine probe and scans the surface of the specimen, point by point, to generate secondary (reflected) electrons from the specimen; these are then detected by a sensor. The image is produced over time as the entire specimen is scanned. There are also scanning transmission electron microscopes (STEM) for analytical use; a scanning beam of electrons penetrates thin samples and provides quantitative determinations of macromolecular masses, and indicates the presence and distribution of the atomic elements in the specimen. The majority of macromolecular (and viral) EM studies are based on TEM.

EM specimens must be solid and thin. Although water is by far the foremost component of living entities, it has long been neglected in EM, as it evaporates in the vacuum of the microscope. Years of research have established many useful procedures for observing biological samples with acceptable structural preservation although their water has been removed (see below). Full preservation of native conformation after complete dehydration nonetheless remains unresolved. Cryo-EM techniques, which began to emerge in the 1980s, consider water as the natural intrinsic component that it actually is. Specimens remain fully hydrated, immobilized in vitreous water (or vitreous ice). Water is cooled very rapidly in

such a way it is solidified without ordering, as it is in hexagonal or cubic ices; the resulting vitreous water is in an amorphous solid state. The specimen is maintained at very low temperature (−170 °C or lower) and the evaporation rate becomes negligible, preserving the hydrated state of the specimen in the microscope [11].

The amount of electron scattering is dependent on specimen density and overall thickness. The specimen must be thin enough to prevent image degradation due to the numerous elastic and inelastic scattering interactions of electrons passing through the specimen [12]. Electrons that interact with the potential field of atomic nuclei usually undergo elastic collisions and are deflected with no loss of energy. They emerge with their wavelength unchanged, but experience an associated phase shift. Elastically scattered and unscattered electrons contribute to image formation. Electrons that interact with the outer electrons of atomic nuclei undergo inelastic collisions and transfer some of their energy to the specimen atoms; this process contributes to image blurring, and is responsible for radiation damage to the specimen. Due to the nonconductive nature of the specimen, these inelastic interactions lead to the formation of many highly reactive ions and free radicals; this results in the breakage of covalent bonds, causing molecular structures to degrade rapidly during observation. Image blurring can be caused by specimen movement, which can include mechanical drift, vibration and/or image thermal drift. These adverse effects can be minimized using imaging strategies in which the electron dose is low [13].

A constant problem with biological specimens is their low contrast, defined as perceptible variation in the intensity between different regions. Two components contribute to contrast, amplitude and phase contrast. Amplitude contrast is due to the corpuscular nature of electrons, and requires use of an objective lens aperture that traps the electrons scattered at large angles (inelastic scattering or high angle elastic scattering). Biological samples scatter electrons weakly (at small angles), and artificial methods are used to enhance scattering properties, such as the introduction of heavy atoms (for example, negative staining). Regions of the specimen that give rise to these trapped electrons appear dark in the image. This mechanism is negligible in unstained specimens. In theory, smaller objective apertures increase the amplitude contrast, but are more prone to astigmatism problems if the aperture is displaced from the center of the beam axis.

Phase contrast is due to the wave-like nature of electrons, and arises from the constructive or destructive interference between the elastically scattered electrons and the unscattered, transmitted electrons. This is the dominant contrast mechanism in cryo-EM of vitrified (unstained) specimens. Elastically scattered electrons experience a phase shift (or difference in path length) as they emerge from the specimen, with no alterations in their amplitude, and this phase shift can be modified. Spherical aberration (inherent to the objective lens of the microscope) and focus setting (directly manipulated by the microscopist) are factors that generate phase contrast. No phase contrast would be seen in an aberration-free microscope when the specimen is imaged exactly in focus (see below).

3.2.2 *Sample Preparation Techniques: Negative Staining, Metal Shadowing, Embedding and Ultramicrotomy*

As important as the development of the electron microscope itself is the development of adequate sample preparation techniques to be used with it. In the late 1950s, detailed images of viruses were obtained by embedding them in a high density (electron-opaque), highly soluble stain, which replicates specimen structure. Heavy metal salts such as phosphotungstic acid, uranyl acetate, and ammonium molybdate have been used successfully as negative stains. Stains stabilize the sample in the harsh conditions necessary for observation, including the hostile atmosphere (high vacuum) and continuous electron irradiation (energy transfer processes). As a general rule, the difficulties associated with biological samples are solved by following four steps: (1) fixation (chemical or physical), (2) dehydration or drying, (3) the introduction of a support (plastic and amorphous carbon layers, resin, metal shadow or vitreous ice), and (4) increase in contrast (heavy, electron-dense elements).

For optimal results, a firm, structureless substrate is essential for specimen support; plastic films such as collodion or formvar stabilized with a thin layer of carbon are appropriate. These films are used as a support in a specimen grid, the electron microscope analog of the glass slide used in light microscopy. EM grids are fine mesh supports (~3 mm in diameter), on which a plastic/carbon film is deposited, and the biological material is adhered for transport and viewing in the electron microscope. Copper is normally used in the fabrication of grids, although other more inert metals such as gold or nickel may be needed.

There are several negative staining procedures; the flotation method is probably the most popular. A plastic/carbon-coated grid is floated on a droplet containing the specimen for 1–3 min to permit adsorption of the specimen. The grids are previously glow-discharged to render the carbon film more hydrophilic and increase specimen adhesion. The grid is then passed through two to three droplets of water, and transferred onto a drop of negative stain for 30–60 s, blotted with filter paper, and air-dried. In principle, the low pH of the negative stain solution, the adsorption of the specimen to the carbon layer, the embedding in a heavy-metal salt cast, as well as the dehydration method can all be expected to severely disrupt the native conformation of the specimen. The technique nonetheless preserves quite well the structure of wide range of specimens, and constitutes a indispensable initial step for higher resolution studies by cryo-EM [14, 15], as resolution is limited by the granularity of the stain to ~20 Å.

The negative staining technique is a very quick and easy method for assessing the homogeneity and quality of many particulate solutions such as purified virus preparations, and requires a minimum of experience and equipment. Limitations are due mainly to structural preservation; stain depicts only the surface features of the specimen and any cavities the stain penetrates, but little or no information is derived about the partially- or unstained internal features. The method is important in

structural virology for fast but unrefined morphological identification. Enveloped viruses are very susceptible to disruption by negative stains and/or air-drying.

Metal shadowing techniques also use heavy metals to enhance the contrast of the low intrinsic electron scattering of biological samples. Shadowing experiments reveal topographic features of specimens by spraying particles of vaporized metal to form a thin coating on whole pre-dried specimens. Fixing and drying of the specimen should preserve the details to be studied; the simplest method is to allow the sample to air-dry, although this technique can result in excessive shrinkage, cracking and/or collapse of fine structures. The freeze-drying method freezes the specimen rapidly while it is still in the aqueous phase. The frozen specimen is transferred to a chamber that maintains a temperature below −80 °C under vacuum, and the solid ice gradually sublimates; preservation is usually better using this process. The freeze-fracture technique replicates fractured surfaces of frozen specimens obtained with a sharp knife; it is especially suited for studies of membrane structure. A variant of this approach is freeze-etching, in which sublimation (etching of ice) is introduced after fracturing.

Platinum or tungsten/tantalum are usually melted and evaporated from a heated electrode, and a second coat of carbon is often used to improve stability of the replica coating. Specimen shadowing can be either unidirectional or rotary. In unidirectional shadowing, the specimen is immobile and is shadowed at an average angle of ~45° in a high vacuum evaporator. This method can be used to measure the height of a specimen from the length of its shadow, and these micrographs are analyzed to obtain topographic maps. In rotary shadowing, the specimen is rotated on a motorized stage and is shadowed from all directions. Metal replicas are released from the underlying biological material by careful chemical treatments such as a hypochlorite solution, and the floating replica is washed thoroughly, carefully fished on fine grids, dried, and then viewed in the microscope.

Embedding and ultramicrotomy of biological samples, including isolated virus particles or a virus-infected tissue or cells, is an elaborate procedure. Specimens are immersed in a matrix that is solidified, then cut into thin slices or sections (50–100 nm thick). The specimen is first chemically fixed to preserve its structure; osmium tetroxide was the first fixative used for EM, but laboratories today use many variations on the fixation process, including potassium permanganate and the standard glutaraldehyde-osmiun tetroxide protocol. The sample is next dehydrated, which consists of the gradual replacement of water with an organic solvent that acts as a "transition solvent" between the aqueous environment of the specimen and the hydrophobic plastic embedding medium; ethanol and acetone are common dehydrating agents. The sample is infiltrated in a resin; in this process, dehydrants are gradually replaced by (viscous) resin monomers such as epoxy mixtures (often called Epon). The resin specimens are transferred into molds or capsules, and placed in an oven in which the epoxy components polymerize to form a solid. There are many types of embedding media with special uses, such as Lowicryl and LR White (acrylic media) used for immunocytochemistry. Finally, the embedded sample is thinly sectioned with a precision (diamond or glass) knife on an

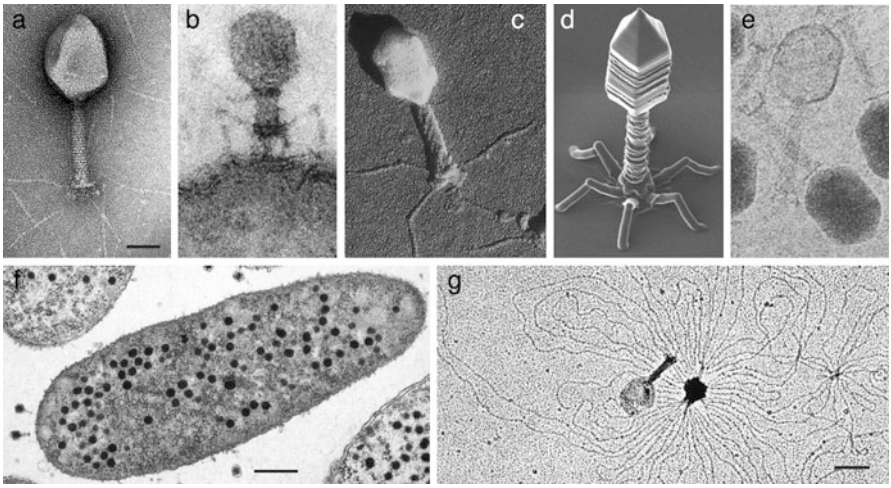


Fig. 3.2 Analysis of bacteriophage T4 by conventional EM and cryo-EM techniques. (**a**) T4 phage negatively stained with 2 % uranyl acetate; darker areas represent local concentrations of stain. The general morphology of the T4 phage, including head, tail, baseplate and fibers can easily be distinguished. The 40.5 Å tail spacing is seen as stacked disks. Note that phage heads are collapsed to different extents, giving the impression of structural heterogeneity due to phage collapse. Bar, 50 nm. (**b**) Thin section showing phage T4 injecting its DNA genome through the bacterial wall; fibers adhere to the outer wall surface and the tail is contracted (Dr. Michel Wurtz, University of Basel, Switzerland; micrograph by J. v.d. Broek). (**c**) Freeze-dried, metal-shadowed phage T4. General morphology is seen as with negative staining, and additional structural details can be inferred: head and tail shadows indicate different heights for these structures; the helical pattern in the tail region is evident in these conditions. (**d**) SEM image of an artificial nano T4 phage fabricated from carbon by focused-ion-beam-induced chemical vapor deposition on a Si surface, about 10× larger than the real virus (micrograph by Reo Kometani and Shinji Matsui, University of Hyogo, Japan). (**e**) Vitrified phage T4; full (*dark*) and empty (*clear*) heads are seen. At this defocus, tail spacing of 40.5 Å is enhanced. (**f**) Thin section through a phage T4-infected *E. coli* bacterium (Dr. Michel Wurtz; micrograph by M. Maeder). Bar, 0.5 μm. (**g**) Air-dried and shadowed broken (by osmotic shock) phage T4 that has released its packaged DNA (Dr. Michel Wurtz; micrograph by J. Meyer). The partially extended DNA indicates high packaging density within the head. Bar, 100 nm

ultramicrotome, sections are picked up on a naked copper grid, and stained with heavy metal salts to enhance contrast.

Bacteriophage T4, of the *Myoviridae* family, is an excellent model that illustrates the complementarity of the information that can be obtained by a variety of conventional TEM procedures (Fig. 3.2). A double-stranded (ds) DNA tailed virus that infects *Escherichia coli*, phage T4 is characterized by a large elongated (prolate icosahedral) capsid (head) (1,150 Å long, 850 Å wide), a tail made of two concentric protein cylinders (1,000 Å long, 210 Å wide), and a complex base plate (460 Å diameter) with six long tail fibers attached to it (1,450 Å long). It is approximately 200 nm long and 80–100 nm wide. The tail is surrounded by a contractile sheath, with helical symmetry (with a pitch of 40.5 Å), which contracts

during infection of the bacterium. dsDNA is densely packaged, ~24 Å interstrand spacings, inside the head [16].

3.2.3 Sample Preparation Techniques: Freeze-Substitution, High Pressure Freezing, and Cryo-Ultramicrotomy

The high vacuum required in the electron microscope column does not allow visualization of living cells in liquid water but, as discussed above, specimen ultrastructure is altered as soon as water is removed. The approaches described thus far increase the stability of dehydrated biological samples by introducing chemical fixatives such as a negative stain agent or osmium tetroxide. As we will see below, the closest approach to the native state is the vitrification of the biological material. This approach is restricted to "thin" structures (viruses are ideal samples), which can be cryo-fixed on a specimen grid by plunge-freezing (rapid immersion in a cryogenic liquid; see below). Thick samples (100–300 μm) can be cryo-fixed by high-pressure freezing.

Freeze-substitution is a low-temperature dehydration method; it can be considered a hybrid technique that bridges the gap between cryo-EM and conventional EM. To ensure good structural preservation, biological samples are cryo-fixed, for example, by high-pressure freezing (see below). They are then freeze-substituted, which consists of dehydration at −90 °C by replacing ice with an organic solvent based on acetone, ethanol or methanol and, if desired, chemical fixatives. After dehydration (or substitution), the temperature is raised to −40 °C to −30 °C and the sample is resin-embedded. Cryo-fixation by high-pressure freezing provides excellent preservation of cellular and molecular architecture for samples between 100 μm and 300 μm thick. This approach has provided realistic views of the crowded eukaryotic cytoplasm, in which protein concentration is ~300 mg/ml.

High-pressure freezing is an approach that permits optimal preservation of biological ultrastructure by cryo-fixation, provided the biological material is not destroyed by water crystallization. This requires high cooling rates (higher than 10,000 K/s) at atmospheric pressure (~1 bar). The native structure is well preserved by conventional freezing (or vitrification) of specimens up to 0.5 μm. Thicker specimens (up to 30 μm) can be frozen without visible ice crystal damage using antifreezing (cryoprotectant) agents, but chemical fixatives are needed; this implies deviation of the native structure and, therefore, loss of high resolution details. Thicker samples (~200 μm) can be adequately frozen under high pressure (~2,000 bars), in which physical properties of water are changed (high pressure is, in fact, a physical cryoprotectant that prevents ice nucleation and crystal growth and increases the depth of vitrification).

Material embedded in vitreous ice is ideal for cutting directly into ultrathin cryo-sections (with a cryo-ultramicrotome using diamond knives) and can be observed in a cryo-electron microscope at liquid nitrogen temperatures, although

cutting-induced deformations and artifacts are a serious problem. This approach is known as cryo-EM of vitreous sections (CEMOVIS). A three-dimensional reconstruction of the specimen in the cryo-section (200 nm thick) can be generated by computerized cryo-ET. CEMOVIS is the best method for preserving whole cells and tissues in their native state, and relatively high resolution is achievable [17].

3.3 Cryo-Electron Microscopy of Viruses

3.3.1 Basic Concepts and General Experimental Design

In cryo-EM of viruses (and other types of biological macromolecular assemblies), the specimen is not dehydrated; instead, it is included in a thin film of water (~0.2 μm or less) that is vitrified [18]. An aliquot of a virus suspension is placed on a grid coated with a holey carbon support film. Excess sample solution is blotted away with filter paper, leaving a very thin film (up to several thousand Å) of virus suspension on the grid, which is rapidly plunged into a cryogen (such as ethane slush) at liquid nitrogen temperature. Freezing is so rapid that water molecules are immobilized in an amorphous state, so-called vitreous or glass-like ice, avoiding physical damage of the specimen by ice crystal formation (Fig. 3.2e). This method provides a natural water-like environment for the virus particle, with no staining or fixing artifacts, and the native structure is preserved with structural integrity near atomic resolution. To avoid devitrification of the sample, the grid bearing the virus sample is maintained at liquid nitrogen temperatures while it is introduced in the microscope with a specimen cryoholder, and for several hours for the recording of micrographs. Alternatively, the grids can be stored indefinitely in liquid nitrogen. Heat transfer between the cryoholder and the microscope can be a serious problem and cause the specimen to vibrate (a result of liquid nitrogen boiling in the cryoholder dewar). Mechanical drift can be another obstacle when the cryoholder is not stably inserted in the column (the specimen must not move by more than a few Å during a 0.5–1 s exposure).

At approximately −170 °C, ice does not sublimate at a significant rate, even in the high vacuum of the electron microscope. Transfer steps must be very rapid to minimize contamination by deposition of hoarfrost (from water vapor), and the cryoholder and the interior of the microscope column are equipped with a cryoshielding device and a twin-blade anti-contaminator, respectively, which closely sandwich the grid. Holey carbon film grids are used as is, or can be glow-discharged, washed with acetone vapor to enhance hydrophilicity, or replaced by continuous non-perforated carbon film grids.

The concentration of the virus suspension is critical for obtaining a thin film with even specimen distribution after blotting; negative staining analysis is used to determine an appropriate range of particle concentration. Another limiting factor is that of the solutes used during virus purification by ultracentrifugation in

gradients such as glycerol, sucrose and cesium chloride. These chemicals will bubble during observation and must be removed by floating the grid on a drop of distilled water (routine for preparing negatively stained samples) or dialysis of the selected virus fractions.

The cryo-negative staining method preserves biological samples in the frozen hydrated state, incorporating negative staining to improve the signal-to-noise ratio of the images. As a contrast agent, ammonium molybdate is normally used (at physiological pH) when this procedure is used for conventional cryo-EM; uranyl formate or acetate is used in the carbon-sandwich method, in which the sample is trapped between two carbon films and is frozen in liquid nitrogen. Unlike air-dried negative staining at room temperature, in the carbon-sandwich method, sample hydration is maintained in all steps (preserving its structural integrity) and specimens are observed at liquid nitrogen temperature [19].

3.3.2 Magnification, Calibration and Minimal Electron Dose

The definition of absolute magnification is necessary, and real magnifications should be calibrated periodically to avoid uncertainty in image processing. The hysteresis of magnetic lenses and changes in specimen height in the objective lens cause small changes in nominal magnification that must be considered in high-resolution studies. Comparisons of three-dimensional reconstruction (3DR) of a virus whose dimensions have been calculated by X-ray analysis is a precise approach to determining the absolute scale; another approach is to mix the test virus with a bacteriophage T4 solution, using the 40.5 Å axial spacing of the bacteriophage tail sheath as an internal magnification standard.

Vitrified specimens are highly sensitive to electron irradiation, and the act of imaging with electrons could eventually destroy the specimen. After insertion of the specimen stage into the microscope, the cryogrid is searched at low magnification ($<$2,000–3,000$\times$) and irradiation levels. An experienced cryomicroscopist is able to select optimal areas by visual inspection; dark areas in the grid are too thick for electron penetration, while bright areas are probably too thin for correct virus particle embedding or can even be dried. Most viruses can be seen directly at low magnification as dark spots, allowing assessment of particle distribution and concentration. These particles are destroyed at high magnification, but can be observed directly for a few seconds as a “bubbling” effect. These sacrificed specimen areas are a convenient form of direct evidence of vitrified specimen quality, assuming that intact adjacent areas in the grid have similar particle concentration and ice thickness.

The strategy for image acquisition is based in this approach, and is done following a relatively “blind” (but not random) procedure, in the sense that the area selected for recording will be seen (and selected or discarded) once the micrograph is developed. Cryo-microscopes are equipped with a low dose system to limit exposure to electrons. The system switches among three states, searching

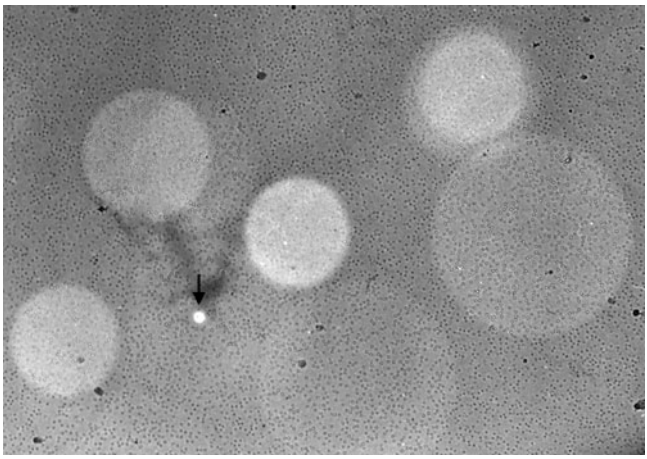


Fig. 3.3 Radiation damage in vitrified bacteriophage T4. In this low magnification field, the smaller circular areas have been sacrificed during focusing; the larger circular areas correspond to recorded regions (the electron dose needed to acquire an image causes no visible effect, but the microscopist exposed the recorded area for longer times to obtain this general view). The bubbling effect can be seen in the particles of some of these exposed areas. The bright area (*arrow*) is literally a hole (made with a focused beam) in the thin vitrified layer that contains the viral particles in suspension

for promising specimen areas at low magnification, focusing at high magnification, and recording at the same high magnification (40,000–60,000×). All of the electrons that pass through the specimen are used to form its image. The image is focused in an adjacent sacrificed area by observing the Fresnel fringes that appear at the edge of radiation-induced bubbles (that first form within the specimen). In this step, astigmatism is corrected and other adjustments are made. Images are recorded with electron doses between 5 $e^-/Å^2$ and 20 $e^-/Å^2$, which produces micrographs of sufficient optical density and signal for subsequent image processing steps (Fig. 3.3).

Imaging on a CCD camera exposes a much smaller sample area, and thus minimizes possible beam-induced specimen movement or charging, similar to the "spot-scan" approach. Inelastic scattering by the sample leads to radiolysis of macromolecules and the embedding medium (vitrified water). The radiolysis products require more space than the original molecules, leading to increase of internal pressure in the sample and ultimately to breakdown (bubbling) of the sample matrix. The pressure increase and the bubbling cause movement in the exposed area of the sample and blurring in the image, thereby attenuating the signal at higher resolution. Inelastic scattering also leads to ejection of electrons from the sample and accumulation of positive charge (sample movement and charging must be minimized).

3.3.3 Imaging Conditions: Contrast Transfer Function

Another drawback in cryo-EM is related to imaging conditions. Images are noisy, because there is no staining agent that enhances the scattering. The amplitude contrast of unstained biological specimens is thus quite low (~10 % or less), and phase contrast is the dominant contrast mechanism, which requires defocusing the microscope objective lens (*i.e.*, images must be deliberately out of focus to enhance contrast). Extensive computer processing is then needed to correct the images and recover structural details.

In a perfect electron microscope, the 2D image recorded on film would be a realistic representation of the projection of the 3D scattering matter in the specimen. As in any other optical system, the image of an object is modulated by the lens characteristics such that each point in the image is the result of a convolution of the corresponding point of the object and a mathematical function, known as the contrast transfer function (CTF). In the case of the electron microscope, the CTF depends on the spherical aberration of the objective lens, the effects of defocusing and inelastic electron scatter, and the partial coherence of the beam. This fundamental concept implies that all TEM images are altered by the CTF and must be corrected for accurate interpretation of the structure of a biological specimen. The CTF is a characteristic function of each microscope (*i.e.*, the spherical aberration coefficient of the objective lens) and the imaging settings (defocus level of the objective lens, beam coherence and accelerating voltage).

By choosing different defocus settings (focal series or pairs are usually taken), specific structural features (or frequencies in the Fourier space) of the image can be accentuated selectively at the expense of others. In appropriate imaging conditions, a contrast-enhancing effect can thus be obtained. CTF-modulated information implies the existence of frequency ranges in which there are no data (zeros in the CTF), regions in which the information has reversed contrast (between the first and the second zero, between the third and fourth zero, and so on), and a fall-off of useful information at higher frequencies (the envelope function), dependent mainly on beam coherence. In the same way that observing a sample by different conventional EM techniques provides complementary data, recording several images at different focal settings to enhance distinct specimen features is a "must-do" (Fig. 3.4).

Cryo-EM micrograph quality can be analyzed by inspection of the CTF zeros or Thon rings in the optical or computed diffraction patterns of images (the sum of the Fourier transform amplitudes is termed the average power spectrum). An image free of aberrations has the signal in the power spectrum as a series of concentric circular rings (zeros or Thon rings) that extends outward from the center of the Fourier transform; the greater the extent of rings, the higher the resolution in the image. Images with drift or astigmatism are easily detected and excluded from further structural analysis (Fig. 3.5).

As good contrast is required for the cryo-micrograph, a compromise should be reached among the factors that modulate contrast. The use of a small objective

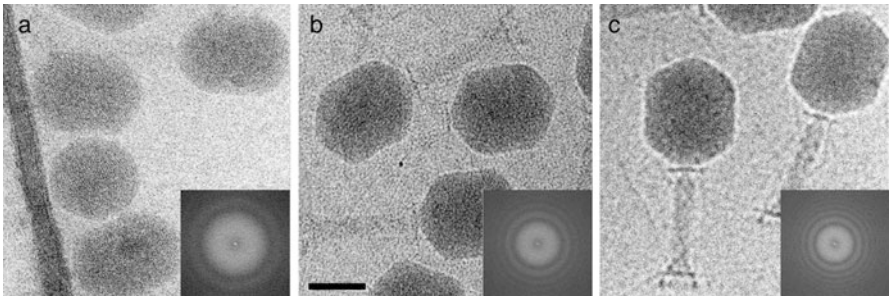


Fig. 3.4 Unstained, frozen, hydrated specimen of phage T4 in vitreous ice, prepared on holey carbon film. Focus series of phage T4 under imaging conditions that enhance dominant spatial frequencies corresponding to (**a**) interstrand spacings between double-stranded DNA filaments (~26 Å) within the head in a relatively close-to-focus image, (**b**) tail sheath spacing (40.5 Å) in a regular defocused image, and (**c**) general morphology in a highly defocused image. Note that detail of the tail is almost lost in the close-to-focus image. Insets show FT (Fourier transform) for each cryomicrograph; the further from focus, the closer the Thon ring is to the FT center. Bar, 50 nm

aperture increases image contrast, but might introduce astigmatism, and lower acceleration voltages increase image contrast, but cause more radiation damage to the specimen. Higher defocus means greater contrast, but also more rapid dampening of the envelope function and many zero transitions of the CTF; CTF of the envelope function is reduced using a brighter beam, such as that generated by FEG. Image contrast can also be improved by additional instrumentation. Energy filters remove inelastically scattered electrons, responsible for much of the noise in micrographs, thus improving the signal-to-noise ratio of the images.

3.4 Cryo-EM Image Processing and Three-Dimensional Reconstruction

Electron microscopes have a large depth of focus compared to virus size. This means that all characteristics at different heights in the virus are in focus simultaneously, and each image recorded in cryo-EM conditions is a 2D projection of 3D scattering density in the specimen. Superposition of all structural information from different levels results in a garbled image. In a single projected view, a specific feature can be derived from any level in the specimen. To reconstruct the native 3D structure, many images must be recorded to obtain different views of the object; viruses tend to be randomly oriented in the vitreous specimen layer [20]. Alternative views can be obtained by tilting the specimen in the microscope. A 10° tilt, which can be accommodated by the anticontaminator, is usually sufficient to give an adequate sampling of orientations. This approach is necessary, for example, to identify the absolute handedness of icosahedral viruses using cryo-EM (see Chap. 2).

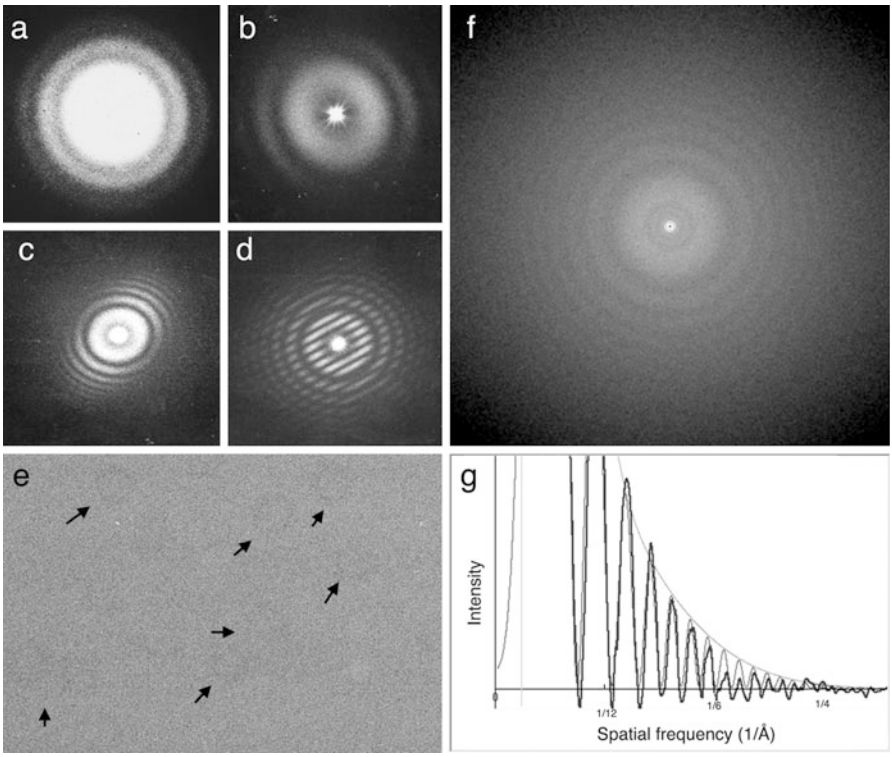


Fig. 3.5 Power spectrum (equivalent to diffraction pattern) analysis of cryomicrographs with distinct aberrations. (**a**) An aberration-free image shows circular concentric Thon rings (zeros). The exact underfocus can be calculated from the first zero position. Signal near the center of the FT corresponds to lower spatial frequencies (low resolution details); signal extending outward from the center corresponds to high spatial frequencies. The last Thon ring indicates the maximum detectable frequency (*i.e.*, resolution) that could be rescued from the 2D projection image. (**b**) Astigmatic migrographs are detected because Thon rings appear elliptical. This aberration can originate from the illumination source and is corrected by the objective lens (an objective aperture that is not well centered can also cause astigmatism). (**c**) Sample drift occurs when resolution is lost preferentially in one direction of the power spectrum. Drift can be thermal (temperature instability in the refrigeration system of the objective lens) or mechanical (stage movement during exposure). Vibrations (acoustic or physical due to liquid nitrogen bubbling in the cryo-stage dewar) limit resolution in all directions. (**d**) Double exposure to determine drift speed. Taking two exposures on the same micrograph, spaced 15–30 s apart, produces a pattern with these typical bands; narrow bands mean larger drift movements and vice versa. (**e**) An excellent image of a rabbit hemorrhagic disease virus solution taken in a FEG 200 kV microscope, in which contrast is very low (*arrows* indicate some viral particles). There is no doubt as to the quality of this micrograph, since the average (**f**) FT calculated from many boxed particle images shows Thon rings to ~4 Å resolution. (**g**) Average power spectrum curve (*thick line*) is used to fit theoretical contrast transfer function (CTF) curves (*thin line*). Underfocus values of 1 μm; first zero of the contrast transfer function at 16 Å

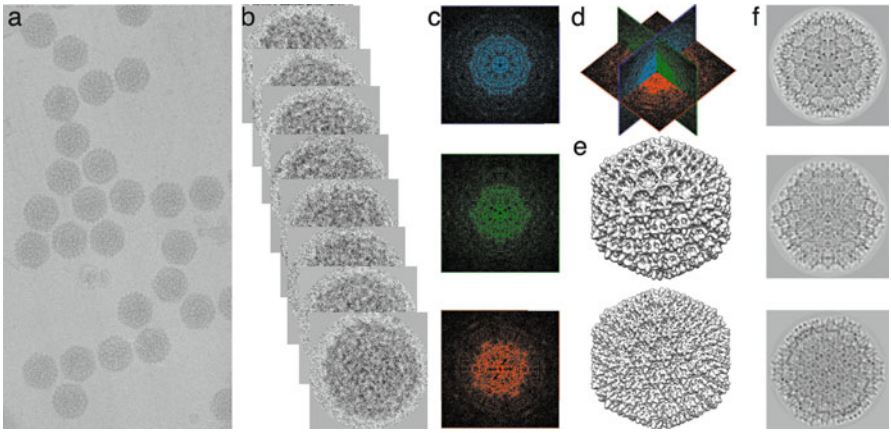


Fig. 3.6 The central section theorem. (**a**) A digitized micrograph containing unstained vitrified viral particles of infectious bursal disease virus (IBDV, a double-stranded RNA virus that infects birds, with a diameter of 650–700 Å). (**b**) Particles are individually selected (*boxed*); (**c**) their 2D FT are calculated from the 2D projections representing different views of the same icosahedral structure. Particle orientation is determined (for example, within 1°) following the projection matching or common lines methods, and the 2D FT are combined in the correct orientation in (**d**) the 3D FT of the virus. (**e**) Accuracy of the initial low resolution 3D map is improved (refined) as the model incorporates more particles and orientation parameters are defined at higher precision in the asymmetric unit. (**f**) Three reprojections from the 3D map are shown

In 1968, DeRosier and Klug found that a 3DR could be recovered from micrographs with 2D images [21], using a mathematical device termed the Fourier transform (FT). Sir Aaron Klug received the Nobel Prize in chemistry in 1982. The FT separates different components of the image into all 2D spatial frequency components (or constituent waves). The FT of each image corresponds to a central plane of the 3D FT of the object density; this is the central section theorem. Distinct projected views fill in different central sections of the transform; the larger the number of views included in the 3D FT, the better the sampling and the 3D object density recovered in Fourier synthesis (Fig. 3.6); many computer programs are available for Fourier analysis and synthesis. In addition to this approach, FT are used for general quality control (evaluating the CTF of each micrograph), correcting the defocus effects (in high resolution studies), and determining particle orientation.

The FT of an image (termed reciprocal or frequency space) is an equivalent, reversible representation of the image (real or Cartesian space). FT provides a mathematical description of the diffraction pattern when imaging an object with a lens (optical or electromagnetic). In the diffraction pattern (the object's FT) produced in the back focal plane of the lens, waves are recombined (positive and negative interference) to form an image. This concept of FT to give a diffraction pattern, followed by Fourier synthesis to produce an image, is central to image processing. In the real-space representation, the parameters that define an image are Cartesian coordinates (x, y, z) and an associated density value. In the FT

representation, these parameters are reciprocal coordinates, an associated amplitude (strength) and an associated phase (which specifies position relative to origin of the wave representing spatial frequency). FT also allows filtering according to spatial frequency. The highest spatial frequencies usually have a large fraction of noise; by eliminating these frequencies in Fourier space and using the inverse FT to convert back to real space, the resulting image is less noisy [22]. The same principles used for 3D image reconstruction are applied for structure determination in X-ray crystallography (Chap. 4).

3.4.1 Digitization

The micrograph must first be scanned with a high precision film scanner to convert it to an array of numbers that represent the darkness on the film; it is then in a form suitable for computer processing. The Shannon sampling theorem states that scanning step size to scan a micrograph should be at least twice as fine as the finest detail to be analyzed (highest frequency or resolution desired), termed the Nyquist sampling. That is, if the resolution desired is 10 Å, 1 pixel should correspond to 5 Å; a 50,000× micrograph scanned with a 7 μm step size corresponds to 1.4 Å/pixel in the specimen; in practice, an image is typically scanned at a scan size of one-fifth to one-third the desired resolution. Another factor to be considered is the limited dynamic range of the cryo-electron micrographs; the scanner should be adjusted to capture this range optimally without image degradation and loss of resolution. This step is omitted if the cryo-electron microscope is equipped with a CCD and digital images are available directly. CCD cameras offer other advantages, such as easy handling, high signal linearity as a function of the number of incoming electrons, and a large dynamic range.

3.4.2 Image Screening

Image quality can be assessed visually by an experienced microscopist to exclude micrographs or CCD frames with high ice contamination, astigmatism, and beam-induced specimen movement. Quantitative analysis is performed by inspection of the CTF rings by calculating the average power spectrum of the micrograph (Fig. 3.4). The CTF is determined by fitting simulated CTF to the observed Thon rings in the calculated power spectrum of the image, which is generally easier for highly defocused images. A CCD camera provides instant feedback for image quality, as FT is calculated dynamically during image acquisition. Once the best micrographs are selected, particles are individually picked (boxed) by manual or automatic methods.

3.4.3 Determination and Refinement of Particle Orientation and Origin

After selection of many thousands of single-particle images (in the best cases), the center of symmetry of each image is established using cross-correlation with a reference image or with the same image rotated 180°. To calculate the 3DR of the object, it must then be determined whether a projection is the front, side or top view, or somewhere in between. Particle orientation determination then begins, that is, the viewing direction relative to the icosahedral symmetry axes. Particle orientations are specified relative to the smallest repeating unit of the icosahedron, termed the asymmetric unit.

The first step of this process requires an initial 3D model as a reference. Many icosahedral virus density maps are available from the Electron Microscopy Data Bank (EMDB: http://www.ebi.ac.uk/pdbe/emdb). The map must be scaled to match the dataset under study, and then low-pass filtered (~35-40 Å resolution) to reduce model bias. A more sophisticated approach involves computing the common lines of a small number of virus images, from which an initial model can be calculated. Due to the inherent difficulties of processing low contrast close-to-focus images, it is often useful to collect focal series with a range of defocus values to facilitate orientation determination. The most time-consuming step when calculating a 3D virus map is the refinement (improvement) of the orientation and center parameters. This analysis involves comparing the particle images with 2D reprojections from a 3D model; projection matching and common lines are the methods most commonly used to determine these parameters [23].

In the projection matching method, the model is projected computationally in all possible directions to generate a set of reference images that are used to match each image with one projection corresponding to the appropriate orientation. The initial low resolution model is updated at the end of each refinement cycle during this iterative process until no further improvement is obtained. Projection matching can be carried out in real or Fourier space.

In the common line method, the central section theorem provides the mathematical foundation for orientation determination of projection images from the same object. The FT of two projection images with different orientations are planes in the 3D FT of the object (a virus) that intersect one another. This intersection line is termed common line and has the same values in each plane for amplitudes and phases (Fig. 3.6).

For an icosahedral virus, a single image with a defined orientation in the asymmetric unit has associated 59 symmetry-related orientations. Each of the 60 equivalent orientations defines the corresponding planes in the 3D FT. The intersection line between any of these planes is known as a self-common line, and there are a maximum of 37 pairs of self-common lines in the FT of a 2D projection image. To estimate individual particle orientation parameters, an exhaustive

search is made for all possible orientations within the asymmetric unit. This method does not require an initial model, but the useful information is limited to the self-common lines.

Cross-common lines arise from the application of icosahedral symmetry between a particle and another particle (with different orientations); there are 60 pairs of cross-common lines between any two images. The orientation angles are initially estimated by assuming an orientation and comparing the agreement of the phases between the common line pairs. All the information in the 2D FT is therefore useful for estimating particle orientation.

3.4.4 Three-Dimensional Reconstruction in Fourier Space and CTF Correction

Considering the central section theorem, and using direct Fourier inversion, a 3D map can be calculated from 2D images [24]. In this method, CTF correction must be applied to the individual particle images before constructing the 3D FT from each 2D FT. It is essential to ensure that sufficient particle orientations have been provided for adequate sampling of Fourier space, since interpolation errors can limit the resolution. This step normally requires intensive memory and computational resources. Although there are other methods for 3D reconstruction of icosahedral viruses, they lie outside our scope here.

The CTF that enables visualization of unstained specimens must be corrected (deconvoluted) to retrieve the high-resolution signal and obtain a 3DR that reliably represents the structure of the specimen. The CTF reverses, removes and attenuates data in frequency ranges in the FT of the image. Precise determination of the Thon ring pattern (CTF zeros or points of minimum contrast) in the power spectrum allows reversing the phases in the appropriate regions of the FT. Multiple images with complementary defocus must be used to fill in the missing data near the zeros. This method is reliable when FEG of high-voltage (200–300 kV) data are used.

3.4.5 Resolution

The resolution of the 3D map depends on the total number of particles, distribution of views within the asymmetric unit, accuracy of center and orientation parameters, the quality of the data (*i.e.*, the radius to which reliable data extend in the FT), and many other factors such as the CTF correction. The resolution in 3DR is less objectively determined than in X-ray crystallography, for which resolution

has come to mean the highest spatial frequency contributing meaningfully to the map [25].

Assessment of the resolution is performed statistically, by splitting the particle set in half and computing two independent maps from each half-set. The two maps (often referred to as even and odd maps) are then correlated as a function of spatial frequency to determine the extent to which the fine structural details have been determined reliably (termed FSC method, Fourier Shell Correlation).

Amplitude scaling is used to restore high-resolution features, using X-ray solution scattering profiles as a reference, or by determining an effective temperature factor (B-factor) (see below). This process of map sharpening scales specific spatial frequencies (resolutions) above others.

Although image variability is a limiting resolution factor, it can also indicate multiple conformations. Conformational states of a macromolecular complex are difficult to classify, as variability must be differentiated from viewing geometry. If these (normally evasive) conformations are assigned to a time-ordered sequence, the resulting density maps can be used to make elegant time-resolved movies of dynamic processes such as virus particle maturation (Chap. 13). As noted above, cryo-EM is the best approach for trapping and visualizing transient structures with a lifetime of a millisecond or longer, as the specimen is rapidly vitrified. Buffer conditions can also be adjusted to extend the lifetime of a transient state; this approach has been used in virus maturation analysis for systems such as herpes simplex virus and several dsDNA bacteriophages (reviewed in [26], an references therein) (see Chap. 13).

3.4.6 Visualization and Structure Representation

Surface rendering of large volume data is very demanding computationally, and can be a rate-limiting step in structural interpretation. For this reason, structural units from the macromolecular complex are normally segmented. Segmentation of individual components permits non-icosahedral averaging of structurally similar components, to enhance signal-to-noise ratio and further improve the resolution of the averaged subunits. For example, the T = 13 capsid of rotavirus allows the nonicosahedral averaging of 13 subunits of VP6 molecules of the asymmetric unit.

In some cases, the structural asymmetric unit boundaries can be established by contouring the map at different levels, based on its compactness and contacts with neighboring densities; homologous models and/or biochemical evidence are generally used. At subnanometer resolutions in the 6–9 Å range, secondary structure elements (SSE) are discernible. The α-helices appear as cylinders of density with diameters of 5–6 Å, and β-sheets as flat densities or planks. An example at this resolution of the final structural model of a spherical virus, the 40 nm-diameter fungal virus *Penicillium chrysogenum* virus (PcV), is shown in Fig. 3.7.

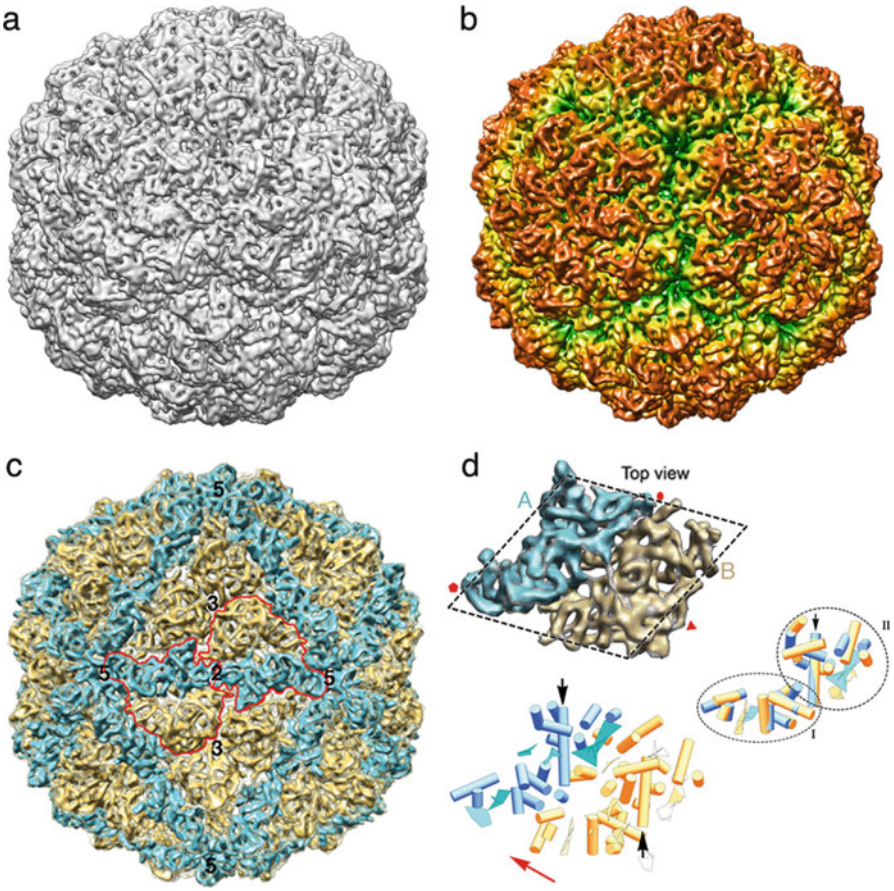


Fig. 3.7 Three-dimensional cryo-EM of the fungal virus Penicillium chrysogenum virus (PcV) at 8.0 Å resolution (based on a FSC threshold of 0.5). (**a**) Surface-shaded representation of the outer surface of the PcV T = 1 (diameter is 400 Å), viewed along a two-fold axis of icosahedral symmetry. Bar, 100 Å. (**b**) As in (a), with the radially color-coded outer surface, which emphasizes the most prominent features as 12 outward-protruding pentamers. (**c**) As in (a); here the asymmetric unit is *cyan* and *yellow*. The asymmetric unit is formed by two ellipsoid-like structures with roughly similar morphology. Boundaries for two asymmetric units are outlined in *red*. Icosahedral symmetry axes are numbered. (**d**) Segmentation of the asymmetric unit, the PcV capsid protein monomer. The *dashed line* highlights the rhomboidal shape, and protein halves A (*cyan*) and B (*yellow*) are indicated (*top left*). At this resolution, secondary structural elements (SSE) of the PcV capsid protein are identified. Cylinders, α-helices; planks, β-sheets. *Black arrows* indicate the ~37 Å-long α-helices of both PcV capsid protein halves. When one of the halves (treated as a rigid body) is superimposed on the other by an ~45-Å translation (*red arrow* indicates translation direction), the relative spatial locations of 13 α-helices and two planar regions are close and require only minor local adjustment. Preserved SSE between both PcV capsid protein halves consist of two domains, domains I and II. This subnanometer resolution analysis indicates that the two PcV capsid protein halves have a comparable structural signature, suggesting ancestral gene duplication (Reproduced/adapted with permission from American Society for Microbiology, ([27], doi:10.1128/JVI.00432-10))

3.4.7 *Hybrid Methods: Combining High-Resolution Structures with Cryo-EM Maps*

As for most research areas, hybrid approaches in structural virology involve multi-disciplinary studies that include a broad set of techniques [28, 29] (see Chap. 7). Here we will refer briefly to the hybridization of atomic structures obtained by X-ray crystallography (see Chap. 4) and by NMR (see Chap. 5) in conjunction with cryo-EM 3DR [30–32]. A more extensive description of combined approaches is included in Chap. 7.

Cryo-EM structural models yield significant information about macromolecular complexes such as viruses (at low to medium resolution), or the folding of individual components (at subnanometer resolution). If more detail is required in the structural model, and higher resolution structural models of individual subunits in the complex are available, the high-resolution structures can be fit ("docked") to the cryo-EM electron density map to obtain a "pseudo-atomic" model of the complex. Fitting tools use rotational and translational searches of a given model within the density map as described in Chap. 7. Such fitting can reveal, for example, interaction interfaces between individual components in atomic detail, providing structural insight into how individual intersubunit interactions might affect complex function. This crucial information cannot, in general, be inferred from the structures of the individual components, and 3D crystals of large functional complexes tend to be very difficult to obtain.

Many studies of large viruses or of the dynamic character of particles that constitute intermediates of virus assembly and function, which are difficult to crystallize, have followed this hybrid approach. Another example is the study of neutralizing monoclonal antibodies and their antigen-binding (Fab) fragments or cell receptors bound to the surface of a virus (see Chaps. 7 and 15). To map the Fab or cell receptor footprint on the virus surface, the antibody-virus complex is studied by cryo-EM and combined with the crystallographic structure of both; this approach was used successfully, for example, for different picornaviruses (rhinovirus, poliovirus and foot-and-mouth disease virus). It is nonetheless possible that the structure of a component will differ when assembled with other components. There are numerous algorithms available for rigid body as well as flexible fitting of components into the larger structure. A quantitative measure of fit accuracy is determined by global and local cross-correlation coefficients.

As the resolution of the cryo-EM map improves, discrepancies in some parts of the individual subunits are evident once fitted into the cryo-EM map; they can be adjusted optimally by permitting flexible fitting of specific domains and/or SSE. Structures in crystals might not always represent conformations related to in situ molecular interactions in physiological conditions. This technique has also been used to validate homology-derived models [33]. Flexible docking can be achieved manually using molecular graphics, but the most-used methods are based on molecular dynamic simulations or on normal mode analysis (Fig. 3.8) (see Chap. 19 for a general description of these theoretical approaches).

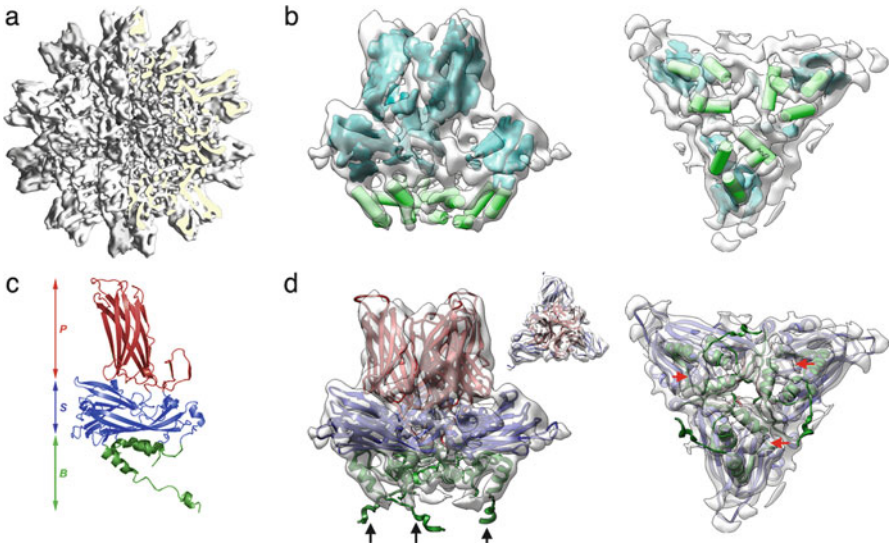


Fig. 3.8 Three-dimensional cryo-EM of T = 1 subviral particles (SVP, ~23 nm in diameter) of IBDV at 7.2-Å resolution. (**a**) Surface-shaded representation of the complete SVP (a quarter has been computationally removed). The most prominent features are 20 VP2 trimeric protrusions of the capsid protein. (**b**) Identification of SSE: green cylinders, α-helices; blue planks, β-sheets. (**c**) The 2.6-Å VP2 X-ray model (PDB entry 2GSY). The VP2 subunit is folded into three domains, termed the projection (*P*), shell (*S*), and base (*B*) domains. Domains S and P are β-barrels. The B domain consists of four N- and C-terminal α-helices that line the inner capsid surface. (**d**) The VP2 trimer X-ray fitted into the corresponding density in the cryo-EM map. The concordance of the two maps is clear in the matching of P and S domain β-barrels. *Arrows* indicate the cut directions to show P and S domain fits (*inset*). Whereas three α-helices fit remarkably well in the cryo-EM structure, a C-terminal α-helix has no corresponding density in the cryo-EM map (*black arrows*), which shows a discernible vacant rod-shaped density adjacent to the C-terminal α-helix (*red arrows*). The proposed movement involves allowed rigid body rotations of the C-α angles in the loop preceding this α-helix. The conformational flexibility of this C-terminal α-helix is critical for IBDV capsid assembly (Reproduced/adapted with permission from American Society for Microbiology ([34], doi:10.1128/JVI.00077-07])

Cryo-EM can also be used as an aid for the atomic resolution of viruses and other biomolecular complexes by X-ray crystallography. Diffraction of crystals of large complexes (such as viruses) requires various phasing methods (see Chap. 4). EM maps are used to solve the phase problem when no phasing by molecular replacement or anomalous scattering is feasible, and heavy atom derivatives do not scatter sufficiently to alter structural factors. A phase extension approach based on molecular envelopes obtained by 3D cryo-EM (or even by conventional negative staining EM) can be used for these crystals (see also Chaps. 4 and 7). If the number of complexes and their positions in the unit cell are known, the envelope specifies those portions of the unit cell whose density is at a constant level. This provides a "solvent flattening" constraint, which with the additional constraint of noncrystallographic symmetry, is imposed iteratively, starting with a low EM resolution

model; the model is progressively improved by introducing diffraction data to higher and higher resolution. This approach has worked well with a number of icosahedral viruses.

3.5 Near-Atomic Resolution of Virus Structures by Cryo-EM

High-resolution 3D cryo-EM is becoming an important structural tool for determining detailed 3D structures of large biological complexes, such as viruses, that are often too large or too heterogeneous to be studied by "conventional" high-resolution methods like X-ray crystallography (Chap. 4) and NMR spectroscopy (Chap. 5) [25, 35–37]. Several recent studies with icosahedral viruses have pushed the limit of single-particle cryo-EM reconstruction to near-atomic resolution (~3.6–4.5 Å). At this resolution, many detailed structural features can be resolved, including deep grooves and pitches of helices, interstrand distance of ~4.4 Å in β-sheets (except for hydrogen-bonded regions), densities for loops and voluminous amino acid side chains, and the zigzag pattern of Cα atoms separated by ~3.8 Å [38]. These structural data can be used as constraints to trace amino acid backbones (Cα model building) using modeling tools. When the X-ray model is available, comparison of the X-ray and cryo-EM structures of the same virus reveals excellent agreement to the level of amino acid side chains.

Several factors have contributed synergistically to this improvement in resolution:

1. *Sample handling*. Structural homogeneity is more important than purity. It is desirable to avoid possible structural damage caused by density gradient centrifugation, since structural heterogeneity (disorder) can be a limiting factor. Icosahedral viruses are regular, more rigid structures than lipid-containing viruses. Dynamics states (*i.e.*, conformational flexibility) also limit resolution.
2. *Cryo-EM imaging*. Improvements in cryo-EM instrumentation such as more stable electron beams and sample holders have reduced specimen drift, a limiting factor in achieving atomic resolution. The high voltage cryo-microscope (FEG, 300 kV) must be optimally aligned so that the electron beam is parallel, with minimal beam tilt. Only those images with visible CTF rings up to 5 Å in their power spectra should be selected (the resolution of the 3DR usually extends beyond the last detectable of Thon ring).
3. *Computational methods for processing noisy cryo-EM images*. Structure refinement is carried out as a two-step, iterative procedure based on projection matching, orientation-origin parameter determination and 3DR by gradually pushing toward higher resolution. High-resolution images are often recorded under close-to-focus conditions, and many thousands of these low-dose phase-contrast images must be used to enhance the high-resolution features. This enormous number of particle images is also related to the fall-off in the Fourier amplitudes at high spatial frequencies (equivalent to high B factor in X-ray

crystallography). Resolution is evaluated by the FSC criterion or by directly evaluating the structural features resolved in the reconstructed map. As in X-ray structures, some regions can have lower resolution than others, probably due to local intrinsic flexibility.

4. *Atomic modeling tools optimized for use with cryo-EM-derived density maps*. Map quality of some protein subunits can be improved by averaging quasi-equivalent subunits, *i.e*., viral capsids with a triangulation (T) number higher than 1 (provided that these subunits are found in similar conformations) (see Chap. 2). An empirical B factor can be estimated by a trial-and-error method that evaluates features such as continuity of backbone and side-chain densities; this can improve high-resolution features and prevent over-sharpening. At this resolution, individual structural components are segmented from the entire capsid and used for model building. At near-atomic resolution (~4 Å), backbone tracing is difficult and error-prone, as some densities are branched and side-chains are limited. Homologous structures obtained by sequence-based comparative modeling can also be used. There are modeling methods adapted to the characteristics of cryo-EM density maps, and modeling tools used in protein crystallography can be adjusted to verify models derived from cryo-EM maps. Evaluation of the atomic models follows the same restrictions used in X-ray crystallography (consistency with protein sequence and reasonable Ramachandran plot, among others). In any case, uncertainties are unavoidable in interpreting maps at 3.5–4.5 Å. Resolved cryo-EM maps are thus a combination of full-atom residue models (SSE-rich regions) and Cα models (regions with less-defined densities).

Near-atomic resolution of virus structures that address important biological questions has been reported for a still-limited number of viruses (reviewed in [24, 25, 35], and references therein). Cytoplasmic polyhedrosis virus was the first example (at 3.88 Å resolution) of a complete *de novo* chain trace from cryo-EM maps. Other examples include the double- and triple-layered rotaviruses, which were unambiguously equivalent to the X-ray model (Fig. 3.9); dsDNA bacteriophage ε15, which is related to bacteriophage HK97; P22, which was resolved in two morphogenetic states; bovine papillomavirus (at 3.6 Å resolution), a dsDNA eukaryotic in which key interactions were described in particle assembly; human adenovirus, which shows how the cement proteins tie together the map structural proteins in the icosahedral capsid; and dsRNA aquareovirus (at 3.3 Å resolution), which showed a priming mechanism for cell entry.

3.6 Reconstructing Viruses Without Imposing Symmetry

Many icosahedral symmetry-based virions present features/structures that do not obey icosahedral symmetry. Examples include the packaging complex, the tail or the core in the capsid interior of dsDNA tailed bacteriophages [40]. These

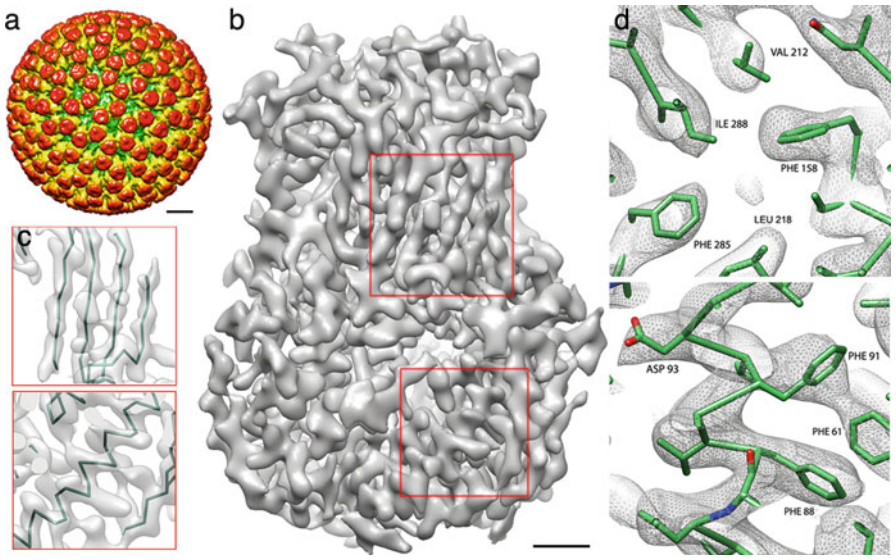


Fig. 3.9 High resolution rotavirus VP6 structure generated by 3D single-particle cryo-EM. A total of 18,125 images of rotavirus double-layered particles (DLP) collected at 300 kV were downloaded from the Grigorieff lab web page at Brandeis University (http://emlab.rose2.brandeis.edu/), and processed using Xmipp software. (**a**) Rotavirus DLP-filtered at ~25 Å. Bar, 100 Å. (**b**) VP6 trimer after 13-fold averaging at ~4 Å resolution. At this resolution, α-helices, β-sheets and densities for loops and bulky amino acid side chains are seen. Bar, 10 Å. (**c**) Close-up of selected VP6 regions included in the squares in (**b**) with the corresponding X-ray structure (PDB 1QHD). Individual β-strands (*top*) and deep grooves and pitches of α-helices (*bottom*) are seen clearly. (**d**) Densities in selected VP6 areas, shown with the atomic model of VP6; bulky side-chain densities match the X-ray structure. These structures are similar to those published by Grigorieff's group [39]

non-icosahedral components can have their own symmetry, which is obscured by the symmetry mismatch with the icosahedral capsid shell (see Chaps. 12 and 17). A map without icosahedral symmetry requires at least 60 times as much data to achieve a similar resolution. Elaborate approaches during the data processing step have enabled structural resolution of complex viruses as a single entity without imposing symmetry, using projection-matching techniques to orient the particles. The asymmetric orientation of a particle is initially determined by a time-consuming search of all possible orientations.

In the case of the T4 phage, the capsid and the tail were reconstructed separately by cryo-EM, after which the atomic models of their component proteins were fitted to determine how the individual parts contribute to the function of this complex molecular machine [41, 42] (see Chap. 11).

3.7 Reconstructing Viruses with Helical Symmetry

The capsids of many plant viruses and bacteriophages have helical symmetry [43]. In addition, the genome of some enveloped animal viruses is surrounded by a helical symmetry-based nucleocapsid; these include filoviruses (Ebola and Marburg viruses), rhabdoviruses (rabies and vesicular stomatitis viruses (VSV)), paramyxoviruses (measles and mumps), orthomyxoviruses (influenza virus) and retroviruses (human immunodeficiency virus, HIV) (see Chap. 2). In certain non-physiological conditions, some capsid proteins of icosahedral viruses may assemble into helical structures, revealing an intrinsic structural polymorphism. Helical viruses and capsids do not crystallize, and many such structures have been determined using X-ray fiber diffraction methods at high resolution. Cryo-EM recently supplied structures at ~10 Å resolution for the VSV virion trunk [44] and 3.3 Å for tobacco mosaic virus (TMV) [45], which allowed the construction of a *de novo* atomic model of TMV in a state that is biologically relevant to its assembly/disassembly.

The redundancy intrinsic to helical assemblies renders them optimal for cryo-EM averaging, since a single image provides all the projection images needed to reconstruct the 3D structure. Helical parameters are determined by measuring the pitch of the helix from layer lines in the FT of the tubular structure. A 2D classification is usually made if structural variability or heterogeneity is detected (or suspected). The 3DR is computed using the iterative helical real space reconstruction (IHRSR) method, a single-particle type approach that does not require long high-ordered filaments. The IHRSR algorithm uses the helical symmetry as a constraint to impose it on asymmetric reconstructions [46].

3.8 Cryo-Electron Tomography of Viruses

3.8.1 Basic Concepts and General Experimental Design

A preparation of icosahedral viruses observed by cryo-EM normally provides multiple views with random orientations of identical particles (without considering structural heterogeneity). Many enveloped viruses are pleomorphic, however; examples include orthomixo-, paramyxo-, retro- and poxviruses (see Chap. 11). Herpesviruses and other enveloped viruses are intermediate forms, with an icosahedral capsid surrounded by a pleomorphic structure made of protein and lipid layers (see Chap. 2). The structural variability of all these viruses means that each particle has a different shape, size or conformation, and 3DR by averaging cryo-EM images as described above is not possible. However, a reconstructed 3D image can be produced by obtaining multiple views of a single viral particle. To do this, the

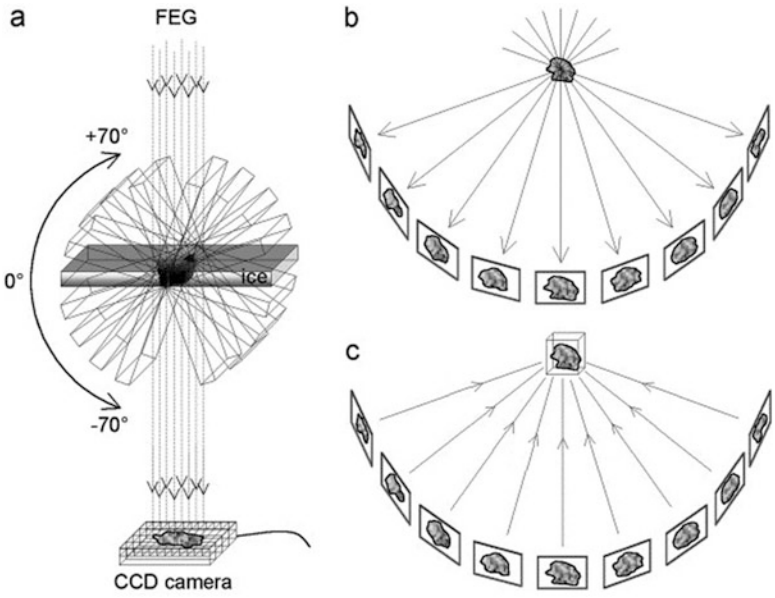


Fig. 3.10 Electron tomography. (**a**) 2D projection images are acquired on a CCD camera while the same specimen area embedded in a vitreous thin layer is tilted incrementally (~1–2°, tilt range ± 70°). (**b**) Diagram illustrating the images projected by a specimen at successive tilt angles. (**c**) After mutual alignment of all these 2D projection images, a density map (tomogram) is reconstructed by a weighted backprojection procedure (Adapted from [47]. Courtesy of W. Baumeister)

specimen is tilted around an axis perpendicular to the electron beam inside the microscope. This type of cryo-EM is termed cryo-ET. Tilting of the specimen provides images of the same field of view (typically more than 100), for example, over a range of ±70° at steps of 1° or 2° (limited by the design of specimen holders and because specimen thickness increases very rapidly at high tilt angles) (Fig. 3.10).

The 3D structure, or tomogram, is calculated as a back-projection in real space; assembly of correctly aligned projections is reverse-projected into 3D space. Cryo-ET has allowed the study of viruses not only in their native physiological state, but also in their natural cell environment (see Chap. 14), thereby providing a bridge between structural studies at the molecular and cellular levels (reviewed in [23, 48, 49], and references therein).

To obtain a detailed, undistorted reconstruction, the tilt series must cover as wide an angular range as possible in as many increments as possible. Tomogram resolution is limited because the range of angles sampled is incomplete and there are no data in some parts of the 3D FT (the missing wedge); in addition, data are missing in the spaces between sampled planes. At the same time, the electron dose must be

kept as low as possible to avoid radiation damage. In principle, a total dose (~100 $e^-/Å^2$) is fractioned over the entire tilt series (~1 $e^-/Å^2$ per image), but the resulting projections are very noisy, making their alignment and CTF correction unreliable. This dose represents less than 5 % of the exposure normally used for single-particle cryo-EM. To increase contrast, images are taken at high defocus values (at least 3–10 μm; compared to 0.5–3 μm for cryo-EM), and resolution is somewhat lower than the first zero of the CTF (in the 4 nm range), as restoration of the true projection is difficult with these low signal-to-noise ratio micrographs. These limitations can be overcome with a double-tilt series; two single-tilt series of the objects are recorded, during which the specimen is rotated by 90° around the beam direction after the first series. The resolution problem is also reduced, if applicable, by image averaging of selected regions (homogeneous subcomponents such as glycoproteins or capsids) of the polymorphic (heterogeneous) viruses.

A cryo-ET setup consists of a high voltage cryo-electron microscope equipped with a FEG, a CCD camera and possibly an energy filter. Automation of the acquisition process is essential for acquiring tomographic data sets in low-dose conditions. High voltage electrons can penetrate thicker samples, which enables imaging of cells or organelles. The interaction of electrons with thick (>200 nm) specimens produces many inelastic scattering events, leading to strong blurring. Energy filters allow imaging of thick specimens and are indispensable for tomography of whole ice-embedded cells or other objects of 0.5–1 μm thickness. Cryo-sectioning is a solution for very thick specimens frozen at high pressure. Without sectioning, only the peripheral areas of flat cells are accessible to cryo-ET.

3DR of an object from its projections involves two steps; first, the micrographs must be aligned to a common coordinate system, and second, the aligned micrographs are merged into a tomogram. Fiducial markers added to the specimen (typically colloidal gold), which appear as high-density dots in the images, are commonly used for alignment, although this can also be done without markers. The most common reconstruction method in cryo-ET is the weighted back-projection, to form a 3DR. Because of uneven sampling in the Fourier space, the low frequencies are artificially enhanced, which requires weighting of projections prior to back-projection to obtain a reliable reconstruction.

The interpretation of tomograms at the ultrastructural level requires decomposition of a tomogram into its structural components, *e.g.*, the segmentation of defined components such as membranes and glycoproteins. Manual assignment of features is commonly used. In addition, to increase the signal-to-noise ratio, so-called denoising algorithms have been developed for analysis and three-dimensional visualization.

Cryo-ET maps of viral complexes at 20–40 Å resolution can be obtained by 3D averaging of different subtomograms from the same particle. Of course, single-particle cryo-EM provides higher resolution maps with an unbiased starting model. Cryo-ET can provide reliable starting models that are then refined by single-particle analysis. In addition to 3D alignments, processing subvolumes can require classification analysis.

3.8.2 *Structure of Pleomorphic Viruses and Capture of Viral Life Cycles Using Cryo-ET*

Cryo-ET has been applied to the visualization of the 3D structures of non-isometric virus particles (reviewed in [50, 51], and references therein); for example, it has been used to study the architecture of influenza virus, herpesvirus, immunodeficiency viruses and vaccinia virus, among other enveloped viruses (Fig. 3.11a).

Influenza virus has two types of protein spikes emerging from its envelope, hemagglutinin (HA) and neuraminidase (NA); there are about 400 densely-packed spikes (85 % HA, 15 % NA) and NA spikes tend to cluster; in addition, HA molecules show conformational changes after fusion with liposomes. Cryo-ET of intact virions of herpes simplex virus 1 (HSV-1) has provided new information about the pleomorphic structures crucial for infection of host cells: tegument (an amorphous protein layer that includes host structures such as actin filaments) and an envelope coated with glycoprotein spikes. The icosahedral $T = 16$ capsid enclosing the viral DNA adopts an eccentric position within the approximately spherical virus. The capsid images were extracted from tomograms, averaged, and compared to the nontegumented capsid, revealing capsid-tegument contacts. The entry portal (a specialized DNA packaging vertex) in different herpesvirus capsids was also visualized in one of the 12 five-fold vertices. An equivalent cryo-ET study to detect asymmetric components inside icosahedral viruses was performed with poliovirus, a ~30 nm icosahedral virus that lacks an envelope; isolate particles were imaged releasing their genome, and RNA exited through channels near the two-fold axes. Cryo-ET of the vaccinia virus identified a complex shell consisting of different layers, in which the inner core membrane incorporates pore-like formations. Envelope glycoproteins (env, gp120) of simian and human immunodeficiency viruses (SIV and HIV) mediate binding to the receptor CD4 to initiate infection; analysis by cryo-ET showed that env trimers change conformation in order to interact with the cell receptor.

Beyond the study of purified viruses, cryo-ET extends its scope to different stages of the viral life cycle and cell pathogenesis (cell-associated state) [57, 58]. Until recently, this information was derived mainly from EM of thin sections of plastic-embedded virus-infected cells; the specimen is chemically fixed, dehydrated in organic solvents, embedded in a resin or plastic, and sectioned. The slices are stained to increase contrast. For cryo-ET, the slices are usually 200–500 nm thick. This differs from the thin (20–100 nm) sections usually used in serial section reconstruction, in which single images of individual thin sections are merged computationally to obtain a 3DR, limiting the resolution in the z-axis to the slice thickness. Specimen thickness is a limiting factor for the preservation of native ultrastructure (unless the high pressure freezing approach is used). Cryo-ET of cells is restricted to unsectioned thin prokaryotic cells and to peripheral thin regions of eukaryotic cells that are flattened, and thus transparent to the electron beam. Analysis of larger cells requires technically challenging cryosectioning.

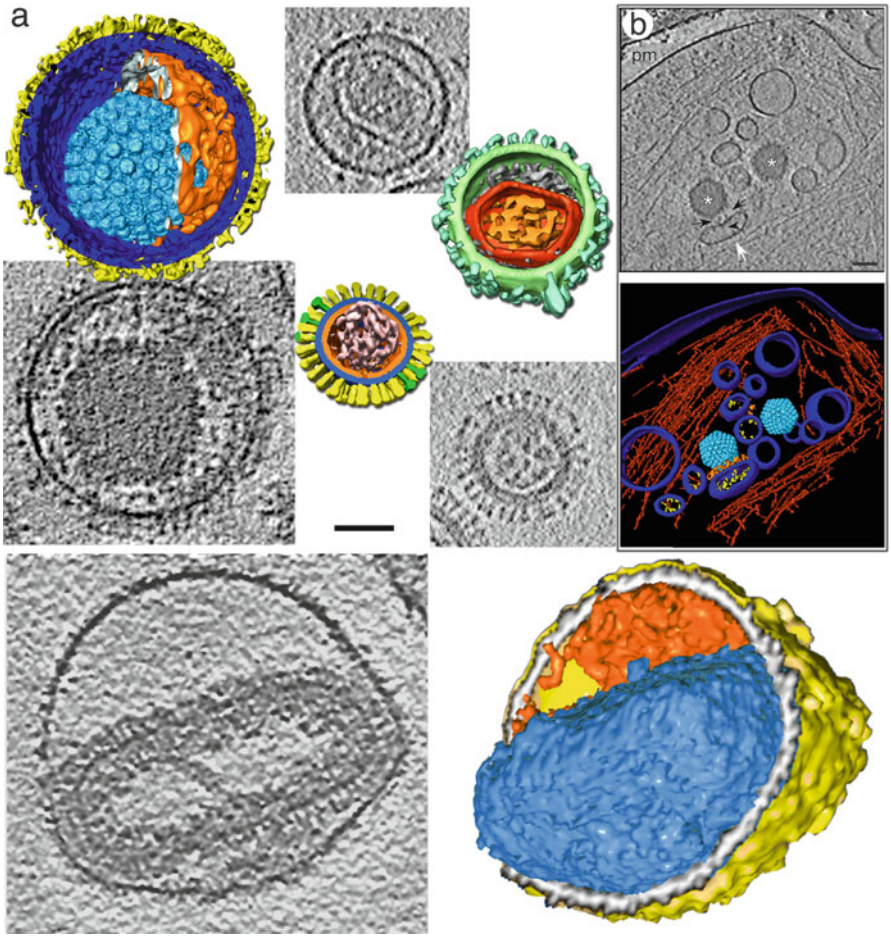


Fig. 3.11 Cryo-ET of pleomorphic viruses and viral life cycles. (**a**) Tomographic reconstructions of HSV-1 (*left*), Rous sarcoma virus (RSV, *right top*), influenza A virus (*right bottom*) and vaccinia virus (VV, *bottom*). Surface-shaded representations of the tomograms are cut to see the virion interior, and a slice of each tomogram is shown. HSV-1 has a T = 16 capsid (*light blue*) buried in the proteinaceous tegument (*orange*), surrounded by a membrane (*dark blue*) studded with glycoprotein spikes (*yellow*) (Adapted from [52]). The RSV virion has a polyhedral capsid (*red*) with an internal density that corresponds to the viral genome (*orange*), surrounded by the matrix protein layer (*grey*); the outermost membrane with the Env spikes is *green* (Adapted from [53]). The genome of influenza A virus is organized as ribonucleoprotein complexes (*pink*); the matrix protein layer (*orange*) is resolved from the lipid bilayer (*blue*); the glycoproteins HA (*yellow*) and NA (*green*) are distinguishable (Adapted from [54]; all three figures courtesy of A.C. Steven). VV mature virion comprises an internal core (*blue*) and the lateral body (*orange*) surrounded by an envelope (Adapted from [55]; courtesy of J.L. Carrascosa). Bar, 50 nm. (**b**) Secondary envelopment of HSV-1 capsids at axon terminals. Slice of the tomogram (*top*) of an intact axon terminal (*asterisks*, capsids; *white arrow*, enveloping vesicle; *arrowhead*, glycoproteins; *black arrows*, tegument; *pm*, plasma membrane). Bar, 100 nm. Surface-shaded representation of the tomogram (*bottom*); color code as in (a); actin is in *red* (Adapted from [56], doi:10.1371/journal.ppat.1002406)

Cryo-ET of the ε15 bacteriophage infecting its bacterial host has allowed visualization of structural changes in the tail machinery, including a tubular density that spans the periplasmic space of the bacterial wall that is the conduit for viral DNA entry into the cell. HSV-1 virions entering synaptosomes (subcellular structures) were caught in action in 3D using cryo-ET; the viral envelope fuses directly with the host membrane, followed by detegumentation. Cryo-ET of frozen-hydrated neurons showed that most egressing capsids were transported independently of the viral envelope; this indicates that progeny herpes viruses are transported along axons as subassemblies and not as complete virions within transport vesicles. Secondary envelopment takes place in axon terminals (Fig. 3.11b). The life cycle of murine gammaherpesvirus, including viral attachment, entry, assembly in virus-induced nuclear inclusion bodies and egress, has been also tracked using cryo-ET. Cryo-ET studies of intact plunge-frozen human cells showed that, during HIV assembly in the native budding sites, the polyprotein Gag lattice is indistinguishable from that of the released immature virion; loss of control of proteolytic maturation led to the formation of aberrant particles.

3.9 Understanding Viruses: Some Major Contributions of Electron Microscopy and Tomography

The first icosahedral virus reconstructions from EM images were made in 1970 from just six negatively stained TBSV (tomato bushy stunt virus) particles at 28 Å resolution, and two of human papillomaviruses at 60 Å [59]. Based on numerous structural virology studies, outstanding progress has been made in the last 40 years. EM imaging has become an indispensable tool in structural virology, and determination of structure is the key to understanding function.

Single-particle cryo-EM has allowed routine analysis of many viral macromolecular complexes in native environments at subnanometer resolutions, in some instances to near-atomic resolution. Cryo-EM is applicable to capsids, genome packaging motors, bacteriophage tails, and other asymmetric supramolecular assemblies (for example, viral polymerase complexes and filamentous ribonucleoprotein complexes). Cryo-EM is best suited to the study of viral complexes too flexible or too scarce to be crystallized for X-ray diffraction studies, or too large for NMR spectroscopy analysis. The most notable feature of cryo-EM is the ability to determine virus structure in its native environment, which permits structural studies of metastable and transient functional intermediate states. The resolution of multiple states will be central to the discovery of molecular mechanisms. 3DR of icosahedral viruses at near-atomic resolution show that atomic models can be derived from cryo-EM images, yielding valuable new information about virus life cycles.

Cryo-ET allows the study of heterogeneous samples such as pleomorphic viruses, and analysis of virus structures in the cellular context, a frontier in

structural virology. Analysis of virus-cell interactions will contribute to the development of strategies to combat a large variety of human, animal and plant viral diseases.

3.10 Perspectives

Although the molecular targets for X-ray crystallography and cryo-EM have begun to overlap as these techniques have been perfected, the synergy between them is clear (see also Chap. 7). Cryo-EM and cryo-ET will continue to provide insights into the structure of large, flexible, biologically relevant macromolecular viral assemblies in increasing, unprecedented detail. Whereas single-particle cryo-EM is a well-established and popular tool for structure analysis, with some anticipated improvements in instrumentation and software, cryo-ET is expected to evolve rapidly in coming years. Introduction of direct electron detectors instead of CCD will improve signal transfer efficiency in cryo-imaging. Likewise, the phase plate cryo-electron microscope has shown promising results in acquiring high-contrast, close-to-focus images with high-resolution information. The need remains for better computational methods for alignment, classification and CTF correction of low-contrast images.

Acknowledgements I thank Daniel Luque for stimulating discussions, José L. Carrascosa and José M. Valpuesta for continuous support, comments and careful reading of the manuscript, Alasdair C. Steven and Benes L. Trus for advice and encouragement and Catherine Mark for editorial help. I am indebted to current and former members of my group (Irene Saugar, Daniel Luque, Nerea Irigoyen, Elena Pascual, Josué Gómez-Blanco, Mariana Castrillo, Ana Correia and Carlos Pérez) and other colleagues for their hard work, skills and enthusiasm that made work possible and enjoyable. This work was supported by grant BFU2011-25902 from the Spanish Ministry of Science and Innovation.

References and Further Reading

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2007) Molecular biology of the cell. Garland Science, New York
2. Nogales E, Grigorieff N (2001) Molecular machines: putting the pieces together. J Cell Biol 152:F1–F10
3. Steven A, Belnap D (2005) Electron microscopy and image processing: an essential tool for structural analysis of macromolecules. Curr Protoc Protein Sci Chapter 17:Unit 17 12
4. Klug A (2010) From virus structure to chromatin: X-ray diffraction to three-dimensional electron microscopy. Annu Rev Biochem 79:1–35
5. Harrison SC (2004) Whither structural biology? Nat Struct Mol Biol 11:12–15
6. Baumeister W, Steven AC (2000) Macromolecular electron microscopy in the era of structural genomics. Trends Biochem Sci 25:624–631
7. Steven AC, Baumeister W (2008) The future is hybrid. J Struct Biol 163:186–195

8. Lasker K, Phillips JL, Russel D, Velazquez-Muriel J, Schneidman-Duhovny D, Tjioe E, Webb B, Schlessinger A, Sali A (2010) Integrative structure modeling of macromolecular assemblies from proteomics data. Mol Cell Proteomics 9:1689–1702
9. Goldsmith CS, Miller SE (2009) Modern uses of electron microscopy for detection of viruses. Clin Microbiol Rev 22:552–563
10. Bozzola JJ, Russell LD (1999) Electron microscopy. Principles and techniques for biologist. Jones and Bartlett Publishers, Boston
11. Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowall AW, Schultz P (1988) Cryo-electron microscopy of vitrified specimens. Q Rev Biophys 21:129–228
12. Amos LA, Henderson R, Unwin PN (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals. Prog Biophys Mol Biol 39:183–231
13. Henderson R (2004) Realizing the potential of electron cryo-microscopy. Q Rev Biophys 37:3–13
14. Harris JR (1997) Negative staining and cryoelectron microscopy: the thin film techniques. BIOS Scientific Publishers Ltd., Oxford
15. Wild P (2008) Electron microscopy of viruses and virus-cell interactions. Methods Cell Biol 88:497–524
16. Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG (2003) Structure and morphogenesis of bacteriophage T4. Cell Mol Life Sci 60:2356–2370
17. Cavalier A, Spehner D, Humbel BM (eds) (2008) Handbook of cryo-preparation methods for electron microscopy. CRC Press, London
18. Baker TS, Olson NH, Fuller SD (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol Mol Biol Rev 63:862–922
19. De Carlo S, Stark H (2010) Cryonegative staining of macromolecular assemblies. Methods Enzymol 481:127–145
20. Crowther RA (2008) The Leeuwenhoek lecture 2006. Microscopy goes cold: frozen viruses reveal their structural secrets. Philos Trans R Soc Lond B Biol Sci 363:2441–2451
21. De Rosier DJ, Klug A (1968) Reconstruction of three dimensional structures from electron micrographs. Nature 217:130–134
22. Moody MF (1990) Image analysis of electron micrographs. In: Hawkes PW, Valdré U (eds) Biophysical electron microscopy. Basis concepts and modern techniques. Academic Press, London, pp 145–288
23. Rochat RH, Chiu W (2012) Cryo-electron microscopy and tomography of virus particles. In: Egelman EH (ed) Comprehensive biophysics, biophysical techniques for structural characterization of macromolecules, vol 1. Academic Press, Oxford, pp 311–340
24. Chang J, Liu X, Rochat RH, Baker ML, Chiu W (2012) Reconstructing virus structures from nanometer to near-atomic resolutions with cryo-electron microscopy and tomography. Adv Exp Med Biol 726:49–90
25. Grigorieff N, Harrison SC (2011) Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr Opin Struct Biol 21:265–273
26. Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF (2005) Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr Opin Struct Biol 15:227–236
27. Luque D, Gonzalez JM, Garriga D, Ghabrial SA, Havens WM, Trus B, Verdaguer N, Carrascosa JL, Castón JR (2010) The T=1 capsid protein of penicillium chrysogenum virus is formed by a repeated helix-rich core indicative of gene duplication. J Virol 84:7256–7266
28. Johnson JE (2008) Multi-disciplinary studies of viruses: the role of structure in shaping the questions and answers. J Struct Biol 163:246–253
29. Russel D, Lasker K, Phillips J, Schneidman-Duhovny D, Velazquez-Muriel JA, Sali A (2009) The structural dynamics of macromolecular processes. Curr Opin Cell Biol 21:97–108
30. Tang L, Johnson JE (2002) Structural biology of viruses by the combination of electron cryomicroscopy and X-ray crystallography. Biochemistry 41:11517–11524

31. Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A (2008) Protein structure fitting and refinement guided by cryo-EM density. Structure 16:295–307
32. Rossmann MG, Morais MC, Leiman PG, Zhang W (2005) Combining X-ray crystallography and electron microscopy. Structure 13:355–362
33. Topf M, Sali A (2005) Combining electron microscopy and comparative protein structure modeling. Curr Opin Struct Biol 15:578–585
34. Luque D, Saugar I, Rodriguez JF, Verdaguer N, Garriga D, Martin CS, Velazquez-Muriel JA, Trus BL, Carrascosa JL, Castón JR (2007) Infectious bursal disease virus capsid assembly and maturation by structural rearrangements of a transient molecular switch. J Virol 81:6869–6878
35. Zhou ZH (2011) Atomic resolution cryo electron microscopy of macromolecular complexes. Adv Protein Chem Struct Biol 82:1–35
36. Zhou ZH (2008) Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Curr Opin Struct Biol 18:218–228
37. Hryc CF, Chen DH, Chiu W (2011) Near-atomic-resolution cryo-EM for molecular virology. Curr Opin Virol 1:110–117
38. Yu X, Jin L, Zhou ZH (2008) 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. Nature 453:415–419
39. Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R, Harrison SC, Grigorieff N (2008) Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc Natl Acad Sci U S A 105:1867–1872
40. Briggs JA, Huiskonen JT, Fernando KV, Gilbert RJ, Scotti P, Butcher SJ, Fuller SD (2005) Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles. J Struct Biol 150:332–339
41. Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov VV, Rossmann MG (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. Nat Struct Mol Biol 12:810–813
42. Aksyuk AA, Rossmann MG (2011) Bacteriophage assembly. Viruses 3:172–203
43. Stubbs G, Kendall A (2012) Helical viruses. Adv Exp Med Biol 726:631–658
44. Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH (2010) Cryo-EM model of the bullet-shaped vesicular stomatitis virus. Science 327:689–693
45. Ge P, Zhou ZH (2011) Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. Proc Natl Acad Sci U S A 108:9637–9642
46. Egelman EH (2007) Single-particle reconstruction from EM images of helical filaments. Curr Opin Struct Biol 17:556–561
47. Grunewald K, Medalia O, Gross A, Steven AC, Baumeister W (2003) Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding. Biophys Chem 100:577–591
48. Lucic V, Forster F, Baumeister W (2005) Structural studies by electron tomography: from cells to molecules. Annu Rev Biochem 74:833–865
49. Cope J, Heumann J, Hoenger A (2011) Cryo-electron tomography for structural characterization of macromolecular complexes. Curr Protoc Protein Sci Chapter 17:Unit17 13
50. Subramaniam S, Bartesaghi A, Liu J, Bennett AE, Sougrat R (2007) Electron tomography of viruses. Curr Opin Struct Biol 17:596–602
51. Grunewald K, Cyrklaff M (2006) Structure of complex viruses and virus-infected cells by electron cryo tomography. Curr Opin Microbiol 9:437–442
52. Grunewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, Steven AC (2003) Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science 302:1396–1398
53. Butan C, Winkler DC, Heymann JB, Craven RC, Steven AC (2008) RSV capsid polymorphism correlates with polymerization efficiency and envelope glycoprotein content: implications that nucleation controls morphogenesis. J Mol Biol 376:1168–1181

54. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci U S A 103:19123–19127
55. Cyrklaff M, Risco C, Fernandez JJ, Jimenez MV, Esteban M, Baumeister W, Carrascosa JL (2005) Cryo-electron tomography of vaccinia virus. Proc Natl Acad Sci U S A 102:2772–2777
56. Ibiricu I, Huiskonen JT, Dohner K, Bradke F, Sodeik B, Grunewald K (2011) Cryo electron tomography of herpes simplex virus during axonal transport and secondary envelopment in primary neurons. PLoS Pathog 7:e1002406
57. Fu CY, Johnson JE (2011) Viral life cycles captured in three-dimensions with electron microscopy tomography. Curr Opin Virol 1:125–133
58. Iwasaki K, Omura T (2010) Electron tomography of the supramolecular structure of virus-infected cells. Curr Opin Struct Biol 20:632–639
59. Crowther RA, Amos LA, Finch JT, De Rosier DJ, Klug A (1970) Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. Nature 226:421–425

Further Reading

Agbandje-McKenna M, McKenna R (eds) (2011) Structural virology. RSC Publishing, Cambridge
Jensen GJ (ed) (2010) Cryo-EM, Part A. Sample preparation and data collection. Methods in Enzymology, vol 481; Cryo-EM, part B. 3-D Reconstruction. Methods in Enzymology, vol 482; Cryo-EM, part C. Analysis, interpretation and case studies. Methods in Enzymology, vol 483. Academic Press
Rossmann MG, Rao VB (eds) (2012) Viral molecular machines. Adv Exp Med Biol, vol 726. Springer, New York

Also especially recommended for further reading are references [3, 4, 11, 23, 25, 26, 27, 35] *listed above*.

Chapter 4
X-Ray Crystallography of Viruses

Nuria Verdaguer, Damià Garriga, and Ignacio Fita

Abstract For about 30 years X-ray crystallography has been by far the most powerful approach for determining virus structures at close to atomic resolutions. Information provided by these studies has deeply and extensively enriched and shaped our vision of the virus world. In turn, the ever increasing complexity and size of the virus structures being investigated have constituted a major driving force for methodological and conceptual developments in X-ray macromolecular crystallography. Landmarks of new virus structures determinations, such as the ones from the first animal viruses or from the first membrane-containing viruses, have often been associated to methodological breakthroughs in X-ray crystallography. In this chapter we present the common ground of proteins and virus crystallography with an emphasis in the peculiarities of virus studies. For example, the solution of the phase problem, a central issue in X-ray diffraction, has benefited enormously from the presence of non-crystallographic symmetry in virus crystals.

Keywords Molecular replacement • Non-crystallographic symmetry • Phase problem • Structural virology • Virus capsid • Virus X-ray crystallography

N. Verdaguer (✉) • I. Fita (✉)
Institut de Biología Molecular de Barcelona (CSIC),
Parc Científic de Barcelona, c/Baldiri i Reixac 10, 08028 Barcelona, Spain
e-mail: nvmcri@ibmb.csic.es; ifrcri@ibmb.csic.es

D. Garriga
Department of Molecular and Cell Biology, Centro Nacional
de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*,
Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_4,

Abbreviations

Ad	Adenovirus
Ad5	Adenovirus type 5
BTV	Bluetongue virus
EM	Electron microscopy
HRV2/HRV14	Human rhinovirus serotypes 2 or 14
MR	Molecular replacement
SBMV	Southern bean mosaic virus
STNV	Satellite tobacco necrosis virus
TBSV	Tomato bushy stunt virus
VLP	Virus-like particle

4.1 Introduction

"Seeing is believing": visualizing the three-dimensional structures of intact virus particles, as well as their constituent proteins and complexes, provides us a tool for a full understanding of their biological function. X-ray crystallography is one of the most powerful approaches for visualizing such macromolecular assemblies at atomic resolution. The first high-resolution pictures of intact virus particles were provided some 30 years ago with the crystal structures of small RNA plant viruses [1–3]. This seminal work was followed by two landmark studies in macromolecular crystallography, the structures of human rhinovirus (Fig. 4.1) and poliovirus [7, 8]. By the end of 1990s, virus crystallography achieved a new breakthrough, the structure of the large core particle (70 nm in diameter) of bluetongue virus (BTV) [9]. The cores of these viruses act as transcriptional machines in infected cells, holding their genomes and transcriptional enzymes within a protective shell so that the viral double-stranded RNA is never revealed to the infected cell, eliminating any chance of cellular response. The BTV structure (Fig. 4.2) showed the assembly of nearly 1,000 protein subunits: the core was made from two principal protein components: a thin skin covering the genome and enzymes, made from 120 copies of a large protein VP3 (triangulation number $T = 2$), which assembles into an icosahedral arrangement *via* conformational switching to yield a pattern of subunits not predicted in the theory of Caspar and Klug (see Chap. 2). This was clothed in 780 copies of the protein VP7 ($T = 13$) in an arrangement that follows the theory of quasi-equivalence.

Enveloped viruses (see Chap. 2) have rarely yielded crystals; however, solubilized protein components of their membranes have been crystallized and studied with great success. Only very regular structures can form single crystals, and in order to study the molecular details of larger and more complex virus

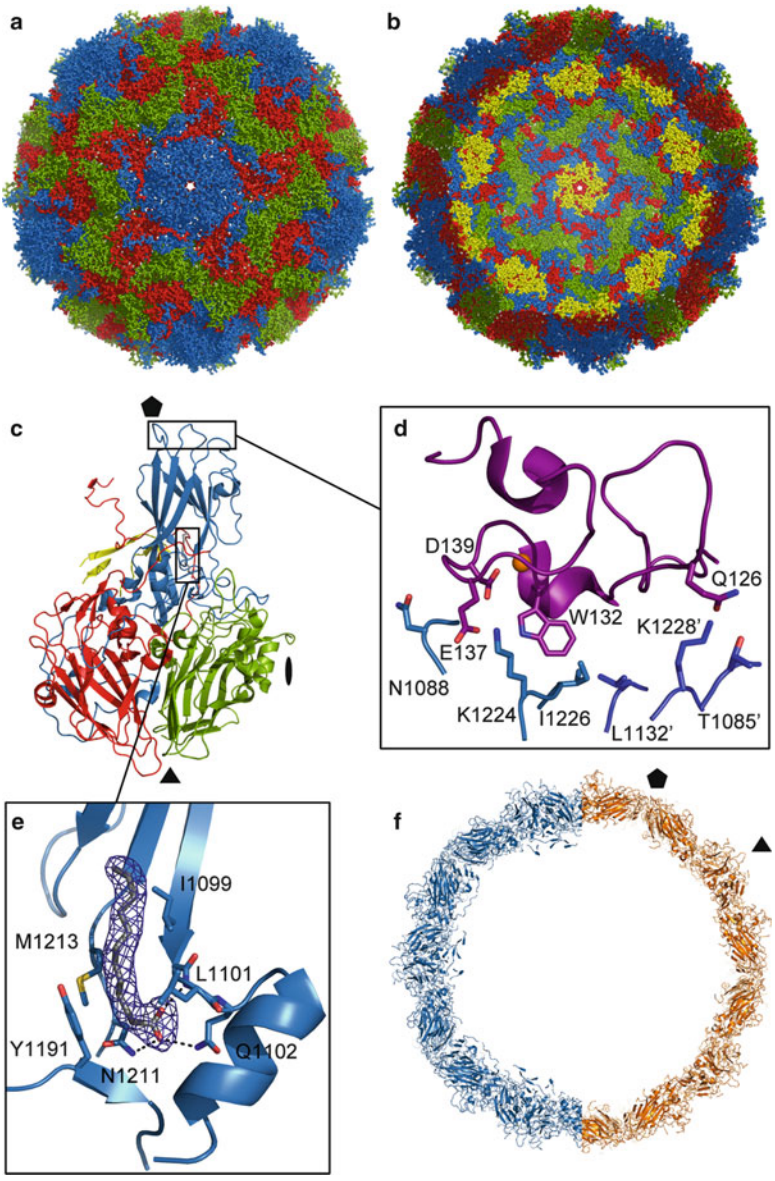


Fig. 4.1 The Human rhinovirus capsid. Representation of the external (**a**) and the internal (**b**) surfaces of the HRV2 capsid (PDB code 1FPN) [4], seen along a fivefold symmetry axis. The capsid, of about 30 nm in diameter, is built by 60 protomers, arranged in a T = 1 icosahedral lattice. Each protomer is composed by four different capsid proteins: VP1 (*colored in blue*), VP2 (*green*), VP3 (*red*) and VP4 (*yellow*). VP1, VP2, and VP3 proteins constitute the capsid shell, whereas VP4 is located at the interior of the capsid, lying on the inner surface, around the fivefold symmetry axes. A prominent feature of the HRV shell is a star-shaped dome on each of the fivefold axes surrounded by a large depression or 'canyon'. (**c**) Ribbon representation of the capsid protomer, colored as in **a** and **b** and with the icosahedral symmetry axes indicated with an *ellipsoid*

particles by X-ray crystallography, it is necessary to dissect them into well defined subunits or substructures. This dissection was originally done with proteases, by disassembly or by isolation of substructures from infected cells. For example, the influenza virus hemagglutinin, the first viral glycoprotein for which atomic details were visualized [10], was obtained from crystals of protein cleaved from the surface of purified virions. In addition, the structure of the adenovirus hexon was obtained from excess unassembled protein derived from adenovirus-infected cells [11]. Since the middle of 1990s, this dissection has more commonly been carried out using recombinant proteins (*i.e.* the gp120 protein from HIV-1; [12]). So far, only one membrane-containing virus X-ray structure has been published: PRD1, possessing an internal membrane, was successfully crystallized, and its structure was solved by X-ray crystallography [13, 14]. Four proteins (MPC, P16, P30 and P31) out of a total of 18 structural proteins identified in PRD1 were resolved in the icosahedrally-ordered capsid. A conformational switch of the MPC N- and C-terminus controls the quasi-equivalent interactions within the pseudo T = 25 lattice.

Today X-ray structural studies of these large assemblies (see also Chaps. 2, 10, 11, 12, 13, 15, 16 and 17) are facilitated by the accumulation of experience in their purification, crystallization and structure determination, as well as by the development of multidisciplinary technologies, including: (i) Crystallization robotics that automate the process; (ii) third generation synchrotron beamlines tailored for data collection from weakly diffracting crystals with large unit cells (see below); and (iii) faster computers and improved software for crystallographic data integration and for structure determination. Current frontiers include defining the architectures of very large assemblies, as exemplified by the 3.5 Å X-ray structure of a complete human adenovirus (Ad) particle [15]. This particle of 150 MDa and an average diameter of 91 nm represents the largest structure solved to date using X-ray

Fig. 4.1 (continued) (twofold axis), a *triangle* (threefold) and a *pentamer* (fivefold). The high resolution crystal structure of a virus is used to gather information for more than the mere spatial disposition of the capsomers; they may also provide the molecular basis of key steps of the virus life cycle. Some examples of this are illustrated in panels (**d**), (**e**) and (**f**), (**d**) Ribbon representation of the structure the HRV2 capsid interacting with the V3 module of VLDL receptor. The cognate receptor for the minor group rhinoviruses (PDB code 1V9U) [5]. One V3 module (*colored in purple*) interacts with two VP1 molecules from neighboring protomers (*colored in different blue tones*). The residues involved in the virus-receptor interaction are represented as sticks and labeled. (**e**) Electron density map (*dark blue wire*) corresponding to the 'pocket factor', a cell-derived fatty acid that often occupies the hydrophobic pocket of VP1, at the bottom of the canyon. The VP1 residues interacting with this pocket factor are represented as sticks. The pocket factor can be displaced by a number of antiviral drugs that are thought to rigidify the capsid, preventing the required structural changes needed for viral genome uncoating during cell infection. (**f**) Comparison of the native HRV2 capsid (*blue*) with the 80S subviral B-particle (*orange*), the end-stage of RNA uncoating (PDB code 3TN9) [6]. Only half of a ~80 Å slab of each capsid shells are represented as to illustrate the expansion in the 80S particle with respect to the native virion. The positions of one threefold and one fivefold axis of the icosahedral particle are indicated with a *triangle* and a *pentamer*, respectively

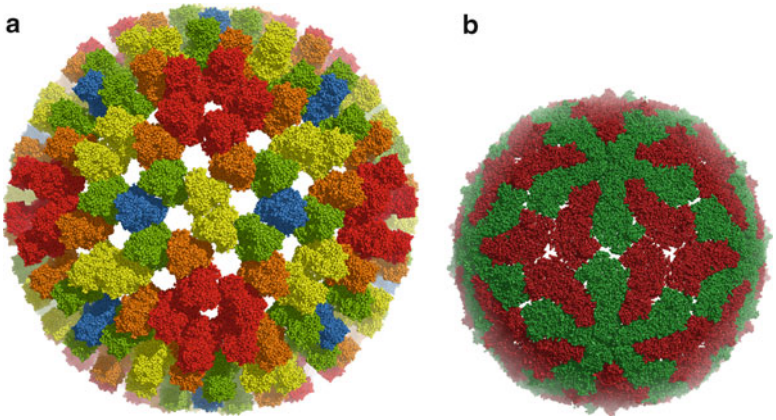


Fig. 4.2 Bluetongue virus core particle structure. Stick representation of the atomic structure of BTV1 virus (PDB code 2BTV) [9], seen along a twofold symmetry axis. The BTV core particle is composed by two differentiated protein shells. The outer shell (**a**) contains 780 copies of VP7 protein, arranged as trimers on a T = 13 quasi-equivalent lattice. In the figure, these VP7 trimers are colored according to their relative position to the symmetry axes (*blue* at the threefold axes, *yellow* at the twofold axes, *red* at the fivefold axes, *green* between the two- and threefold axes trimers, *orange* between the two- and fivefold axes trimers). The inner shell (**b**) consists in a thin subcore generated by 120 copies of VP3 protein organized in a T = 2 lattice. The two structural conformations of the VP3 protein are shown in *dark red* and *dark green*. Both layers of the core are represented at scale

crystallography (Fig. 4.3). Furthermore, comparative analyses of multiple structures of these assemblies, captured in different conformational states, greatly help to precisely define how they function and are regulated. The growing number of structures of virus-receptor complexes [16] (Fig. 4.1b) (see Chap. 15), mature and immature capsid proteins [17] (see Chap. 13), membrane fusion intermediates [18, 19] (see Chap. 16) or RNA uncoating intermediates [6, 20] (Fig. 4.1f) (see Chap. 15) illustrate some examples. It is now clear that solving the structure of a new virus at high resolution typically signifies the beginning, not the end, of a modern structural virology study.

4.2 Basic Concepts and General Experimental Design

4.2.1 Basic Concepts

We cannot form images of things that are much smaller than the wavelength of light we are using. X-rays are in the order of atom diameters and bond lengths, allowing atoms in a molecule or molecular complex to be individually resolved. However, X-ray scattering from a single molecule would be unimaginably weak. In contrast, crystals arrange a huge number of molecules (*e.g.*, virus particles) in the same

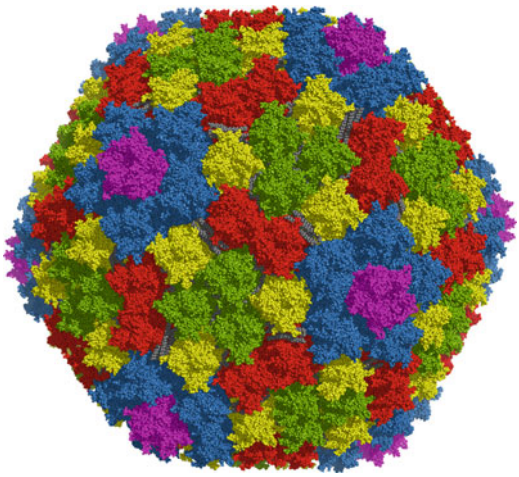


Fig. 4.3 Structure of Human adenovirus. Stick representation of the crystal structure of Ad5 (PDB code 1VSZ) [15]. The adenovirus capsid is mainly composed of hexon protein, with 720 subunits arranged as 240 trimers on a pseudo-T = 25 icosahedral lattice (12 hexon trimers per icosahedral facet). In the figure, the hexon trimers are colored according to their relative position to the symmetry axes: *cyan* around the fivefold axes, *green* around the threefold axes, *red* around the twofold axes and *yellow* between the two- and fivefold axes. Each of the 20 vertices of the capsid is occupied by five copies of the penton base protein (*colored in pink*), in which the Ad fiber shafts are inserted (not shown in the figure). Additional proteins (proteins IIIa, VI, VIII and IX, the so-called cement proteins, represented as *grey balls* in the figure) mediate and stabilize the interactions between neighboring hexon trimers

orientation and scattered waves add up in phase, increasing the signal to a level which can be measured.

Figure 4.4 shows two examples of packing of a virus in two different crystal types. The lattice is defined as a construct dividing space into regular translationally periodic units. The combination of the unit lattice with molecular motifs (*e.g.* a virus particle) generates the unit cell of the crystal that is built from translationally stacked unit cells. In addition, the molecules contained in each unit cell can be related by internal symmetry, generating a unit cell packed with multiple symmetry equivalent copies of the molecule (Fig. 4.4b). The number of different ways to pack molecules in space is limited. For asymmetric chiral molecules (such as proteins), there are only 65 discrete and distinct ways to assemble three-dimensional periodic crystals through combination of translational and rotational symmetry. These 65 types of arrangements form 65 space groups. Their symmetry properties and the rules for constructing each crystal structure are described in the International Tables for Crystallography, Volume A [21].

The asymmetric unit of a unit cell contains all the necessary information to generate the complete unit cell of a crystal structure by applying its symmetry operations to the asymmetric unit. Besides the strict crystallographic symmetry between molecules, additional non-crystallographic symmetry can locally exist

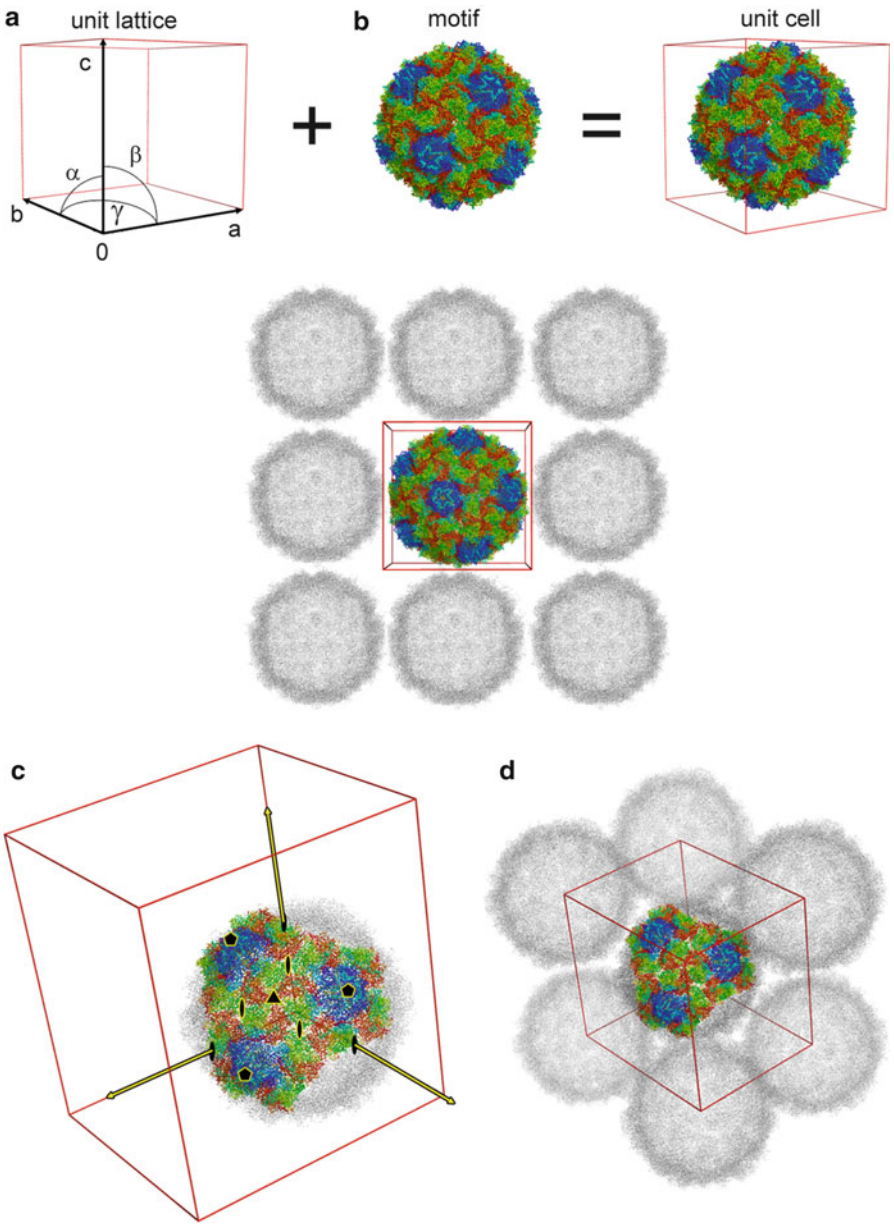


Fig. 4.4 Assembly of virus crystals. (**a**) The crystal unit cell is built by a motif (in this case, a whole virus) that fills the unit lattice. The unit lattice is defined by three basic vectors [**a**,**b**,**c**] that extend from a common origin [0] in a right handed system. The angle between **a** and **b** is γ, the angle between **b** and **c** is α and the angle between **a** and **c** is β. The stacking of multiple unit cells will produce a crystal. (**b**) Crystal packing of an icosahedral virus, in which the whole virus is the asymmetric unit (*colored*). The other viruses of the crystal (*grey*) can be obtained by translation of the unit cell. For clarity, only one two-dimensional layer of the crystal is shown. (**c**) An icosahedral capsid has 12 fivefold symmetry axes (indicated with pentamers, in the figure), 20 threefold

between molecules in the asymmetric unit. In virus crystals, the molecular symmetry results in most cases from the combination of local and crystallographic symmetry (Fig. 4.4c).

In a basic diffraction experiment, a single molecular crystal is placed into a finely focused X-ray beam and the diffraction images are recorded in a detector. The electron density representing the atomic structure of the molecules in the crystal is reconstructed by Fourier Methods from the diffraction data and an atomic model of the structure is built into the electron density (Fig. 4.5). It is important to note that the diffraction experiment is fundamentally different from imaging with an optical microscope. The crucial difference is that visible light scattered from objects can be focused through refractive lenses to create a magnified image of the object. In contrast, as no refractive lenses are available for X-rays, the X-ray diffraction patterns do not provide a direct image of the molecular structure. The electron density of the scattering molecular structure must be reconstructed by Fourier transform techniques. For this type of reconstruction, from reciprocal diffraction space back into the direct molecular space, two terms are required: (i) the structure factor amplitudes, directly accessible in the form of the square root of the intensities of each measured diffraction spot or reflection and (ii) the relative phase angle of each reflection. These phase angles are not directly accessible and must be supplied by additional phasing experiments. The absence of directly accessible phases constitutes the so-called phase problem in macromolecular crystallography (see Sect. 4.6).

At the end of the chapter (Section on *Further reading*) three reference books are provided for an extended description of the general principles and techniques in X-ray crystallography of biological molecules. The first two books are aimed at non-practitioners; the last one contains a more in-depth treatment.

4.2.2 General Experimental Design

The general approach to structure determination of viral particles (and other biomolecules and molecular complexes) by X-ray crystallography is schematically shown in Fig. 4.5. The procedure involves the following basic steps: (i) preparation and purification of virus particle samples; (ii) crystallization and mounting the virus

Fig. 4.4 (continued) symmetry axes (*triangles*) and 30 twofold symmetry axes (*ellipsoids*), relating the quasi-identical subunits that build up the virus capsid. When this virus is fitted in a crystal lattice, it may happen that some of the icosahedral symmetry axes coincide with the symmetry axes defined by the unit cell (*yellow arrows*); in this case, only part of the virus forms part of the crystal asymmetric unit (*colored*). The rest of the virus particle (*grey*) can be obtained applying the crystal symmetry operators to the asymmetric unit. (**d**) The crystal packing of an icosahedral virus, in which one forth of the virus forms the crystal asymmetric unit

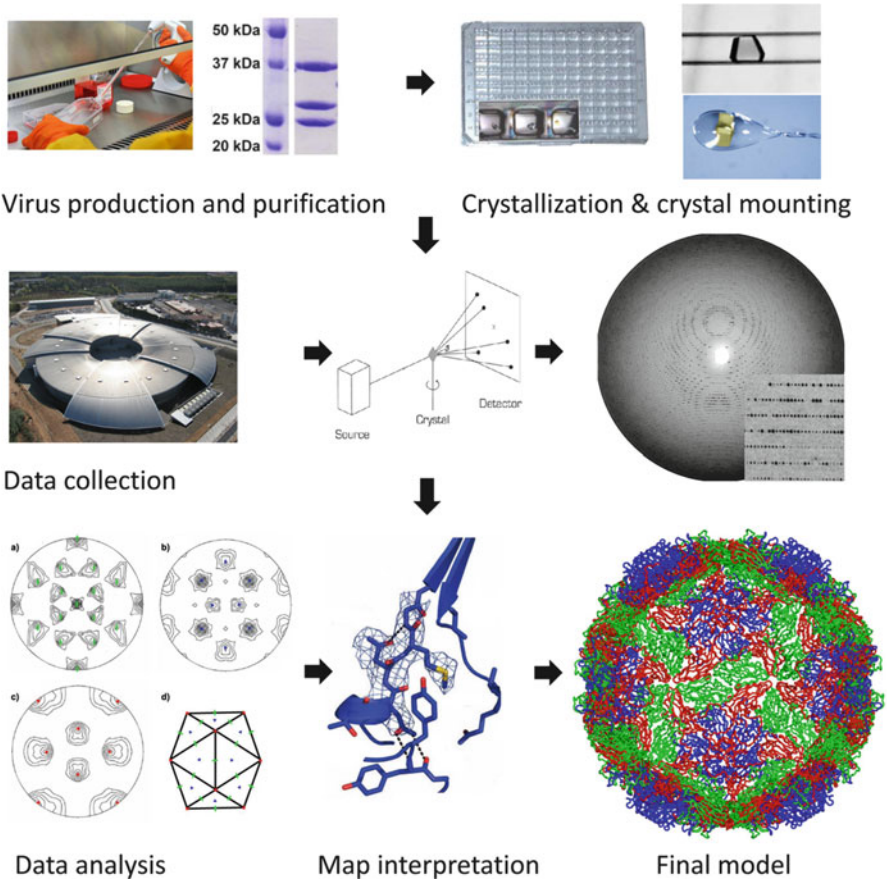


Fig. 4.5 Virus structure determination by X-ray crystallography. The solution of a virus structure involves the following steps: (i) preparation and purification of virus samples, (ii) crystallization and mounting the virus crystals, (iii) measurement of the diffraction data, (iv) Phase calculation usually by molecular replacement and, (v) map interpretation and model building

particle crystals; (iii) measurement of the diffraction data; (iv) phase calculation, usually by molecular replacement (MR); and (v) map interpretation and model building. These steps will be briefly described in the following sections in this chapter (Sects. 4.3, 4.4, 4.5, 4.6 and 4.7).

The first limitation for an X-ray experiment is the requirement of well-ordered three-dimensional crystals. This implies that we cannot determine the structure of a macromolecular complex that does not have a single, well-defined, structure or that is difficult to obtain in a pure and stable state. Furthermore, the structure determination of an intact virus particle entails a significant scaling up in most respects compared to a traditional crystallographic analysis of a single protein. At a fixed resolution, the number of reflection spots to be measured is proportional to the volume of the asymmetric unit of the crystal. For a virus particle, there are several

hundred times as many data to be measured than for an isolated protein. Not only are there far more X-ray reflections to be measured, but these reflections will each be, on average, several hundred times weaker. Therefore, virus crystallography makes considerable demands on the qualities of X-ray sources, the performance and sensitivity of X-ray detectors and computer hardware and software [22].

Having a well-diffracting single crystal is a first step, though not necessarily sufficient, to solve the structure. As mentioned in Sect. 4.2.1, X-ray detectors are able to measure the position and intensity of reflections without recording the phase information. Therefore, in order to compute the Fourier transform, to go from the diffraction pattern to the electron density map of the structure, we need to obtain information about the phases. In virus crystallography, molecular replacement is now the method most frequently used to estimate the phases, needed to solve crystal structures (see Sect. 4.6). Spherical viruses are made from protein shells with icosahedral (532) symmetry (Chap. 2). According to the theory of quasi-equivalence, the icosahedral shell is composed of $60 \times T$ chemically identical subunits which can then, in principle, make similar interactions to form the capsid (see Chap. 2). The presence of icosahedral symmetry leads to the presence of non-crystallographic redundancy (with a minimum of fivefold non-crystallographic symmetry, for those viruses that crystallize with the viral capsid lying on a point of 23 crystallographic symmetry). This redundancy is of enormous importance in a crystallographic analysis, providing extremely powerful constraints that facilitate many aspects of the analysis [22] (Sect. 4.6).

4.3 Production and Purification of Viral Particles and Proteins for Structural Studies

4.3.1 Growth and Purification of Intact Viruses and Viral Capsids

The first step in the route to solving the structure of a viral particle by X-ray crystallography (Fig. 4.5), and a prerequisite for crystallization of the particle, is the availability of sufficient amounts of the sample in purified form. For virus particles, the purification from native sources often relies on classical biochemical protein purification procedures. Virus yields vary and for animal viruses grown in a monolayer cell culture, a vast number of roller bottles or large culture plates may be required. A key step in purification is usually ultracentrifugation using sucrose and/or CsCl gradients. In addition, chromatographic FPLC techniques are now available to separate high-molecular-weight species. Many virus particles may be readily concentrated using 100-kDa microconcentrators or precipitated by ultracentrifugation. Standard spectrophotometric concentration measurements should be corrected for the nucleic acid content; an approximate formula is: Virus particle concentration $= (1.55\ A_{280} – 0.76\ A_{260}) \times$ dilution [23].

With a macromolecular assembly, there is a possibility of crystallizing a substructure such as an assembly intermediate (see Chaps. 10 and 11). Homogeneity and integrity of the sample used for crystallization and of the crystallized molecule should be carefully analyzed. The crystallization of intact virus particles can be verified using SDS-PAGE to demonstrate the presence of the correct complement of proteins, using tests to show recovery of infectivity from dissolved crystals and by electron microscopic analysis of the crystals.

4.3.2 Expression and Purification of Recombinant Viral Proteins

Low yield virus production and sample heterogeneity sometimes impede purification of native virions with the quality and in the quantity required for structural studies. The developments in molecular biology techniques that allow cloning the viral genomes or capsid protein genes for overexpression and assembly of recombinant virus-like particles (VLPs) in heterologous systems are instrumental for the structural studies of theses viral assemblies. Recombinant techniques can further provide the means to modify complexes at the gene level, resulting in alteration of the protein complex subunits to meet the high quality requirements for structural analyses, as well as for site-directed structure-function studies. A successful example in the application of these techniques was achieved in the X-ray structure determination of human adenovirus type 5 (Ad5). High quality crystals were obtained of a genetically modified virus, in which short penton fibers from another serotype have replaced those of the wild-type Ad5 [15].

Molecular cloning, protein expression and crystallization are the major preparatory steps of a structural study. All three processes are experimentally driven and its outcome cannot be predicted. A parallel approach pursuing multiple variants of the protein target significantly increases the chances of success. A number of high throughput initiatives have been developed using affordable methods and equipment to automate molecular cloning of expression constructs, recombinant expression screenings and purification. The advantages of automation are obvious: procedures can be carried out in parallel and scaled-up accordingly to process large amounts of samples at reasonable cost. The methodology is also available for overexpression of multicomponent protein complexes, either in prokatyotic or eucaryotic cells [24].

4.4 Crystallization

The biochemical and biophysical process of crystallization is still a largely empirical process. Crystals grow from an aqueous protein solution, when the solution is brought into supersaturation (Fig. 4.6). Supersaturation is achieved by varying concentrations of precipitant, protein, additives, and other parameters affecting solubility, such as pH and temperature. Crystallization proceeds in two phases,

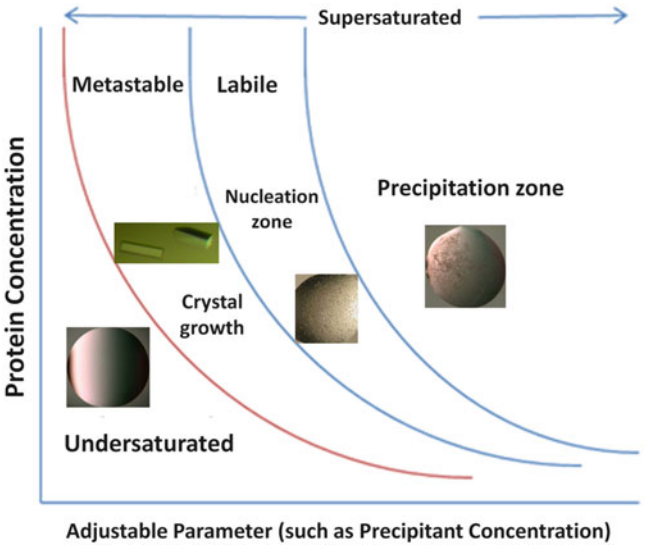


Fig. 4.6 Crystallization Phase Diagram. A schematic drawing of a protein crystallization phase diagram based on two of the most commonly varied parameters, protein and precipitant concentrations. The figure schematically illustrates four areas: (i) an area of very high supersaturation where the protein will precipitate; (ii) an area of moderate supersaturation where spontaneous nucleation takes place; (iii) an area of lower supersaturation, just below the nucleation zone where crystals are stable and may grow. This area, known as metastable zone, is thought to contain the best conditions for growth of large and well ordered crystal: (iv) an undersaturated area where the protein is fully dissolved and will never crystallize

nucleation and growth. Nucleation is the conception stage of crystallization during which a molecular aggregate of critical size forms, allowing for subsequent crystal growth. Once the crystal nucleus has formed, crystal growth follows spontaneously (see [25] for a complete view).

The crystallization process can be effectively illustrated using a phase diagram (Fig. 4.6). The phase diagram indicates which state is stable under different chemical and physical conditions. It provides means of quantifying the influence of parameters such as the concentrations of protein, precipitants, additives, etc. on the production of crystals, allowing the design of crystal optimization experiments. Crystallization conditions for viruses and other complex assemblies are similar to those of single proteins.

4.4.1 Classical Methods

There are four major techniques for growing crystals of biological macromolecules. These techniques have been in use for over 40 years and include: batch crystallization, dialysis, liquid-liquid diffusion (also known as free interface diffusion), and

vapor diffusion [25]. The four methods work well, the later being the most extended technique for viral particle crystallization. The typical time of crystal growth, using any one of these techniques, is 1-3 weeks. However the time for obtaining crystals can vary from several hours to as long as a year. Standard virus concentrations for the first crystallization trials range around 10–20 mg/ml; however, if the solubility of the virus is limited to less than this, it is still worth trying at whatever concentration can be achieved.

The method of vapor diffusion begins with an aqueous crystallization drop containing the macromolecule and the crystallization agents at a concentration lower that that required for the formation of crystals. This drop (typically from 0.5 μl to 10 μl) is equilibrated against a reservoir (0.5–25 ml) containing a solution with crystallization agents at different concentrations than in the crystallization drop. Usually, the crystallization cocktail is used as reservoir solution but it is not a requirement, the reservoir solution and the crystallization cocktail can be different chemical species [26]. Equilibration proceeds by diffusion of the volatile species, typically water, until vapor-pressure equilibrium is reached between the experiment drop and the reservoir. As a consequence of the equilibration process, there is an increase in the relative concentration of solutes in the crystallization drop. By decreasing the amount of water in the drop, the reservoir acts to gradually concentrate the protein and other chemicals in the crystallization drop. This gradual change and increase in the concentration may lead to supersaturation of the macromolecule and subsequent crystallization. Three commonly used vapor diffusion methods are; hanging drop, where the drop is suspended over a reservoir from the underside of a hydrophobic surface; sitting drop, where the drops sits in a small concave surface; and sandwich drop, where the drop makes contact with two narrowly separated surfaces.

The first step of finding initial crystallization conditions for a new macromolecule is to set up screening trials, exposing the macromolecule to numerous different chemical agents in different drops in order to find ‘hits’ or ‘leads’ that point to conditions that may be conducive to crystallization. Crystals, crystalline precipitate and phase separation are all considered leads that are worth pursuing.

One of the main phenomena preventing crystallization is particle aggregation. Electron microscopy is a useful tool for checking this, in addition to methods such as dynamic light scattering. If aggregation proves to be a problem, it may be relieved by the addition of detergents. Other useful additives include divalent cations known to interact with the capsid. With viruses, highly ordered crystals are often obtained once the optimum crystallization conditions are found. This may be attributable to their isometric nature and reflects the perfection with which these systems assemble. There is a tendency toward high symmetry space groups, often with some of the capsid symmetry axes coinciding with those of the crystal [22].

4.4.2 New Approaches

Today structural studies of large molecular assemblies such as viruses are facilitated by the development of the robotic crystallization technology. The use

of robotic nanoliter screenings allows exploring a vast number of different conditions with a small amount of sample [27]. High throughput technologies render the initial crystallization screening process, an often repetitive process utilizing a similar set of cocktail conditions, less labor-intensive and sample-consuming. Many stages of the procedure are automated, starting with the mixing of concentrated chemical stock solutions into a set of crystallization cocktails, to dispensing the macromolecule plus precipitant mixture droplets into plates, sealing the plates, monitoring, scoring the outcomes with the use of image recognition software and even suggesting new experiments based on those outcomes.

4.5 Data Collection and Processing

Once adequate crystals of the viral particle are available, they are mounted in an appropriate holder and irradiated with X-rays (Fig. 4.5). Macromolecular crystals commonly suffer rapid radiation damage during room temperature X-ray data collection. Therefore, data are now routinely collected with the sample held at around 100 K, usually resulting in higher resolution and better quality data. In addition to the cryo-protection techniques, other sophisticated post crystallization treatments, as crystal dehydration, resulted crucial for the structure determination of large macromolecular assemblies.

4.5.1 Cryo-Crystallography

Macromolecular crystallographic data are nowadays routinely collected at cryogenic temperatures (100 K). The key advantage of this technique is the enormous reduction in secondary radiation damage to the sample during X-ray data collection compared to data collection at room temperature. This is because, at cryo-temperatures, most of the free radicals produced by the incident radiation have much lower diffusion rates so they do not travel as far through the crystal, thus avoiding as much damage. The prolonged crystal life-times usually result in better quality and higher resolution data. Nevertheless, even in samples that are cryo-cooled, radiation damage is often observed at undulator synchrotron beamlines. It is a limiting problem in the optimum use of these high brilliant beams, which are routinely attenuated or defocused.

Currently, mounting a crystal for cryo-crystallography involves suspending it by surface tension in cryo-buffer within a thin fiber loop (Fig. 4.5) and then flash-cooling it in gaseous or liquid cryogen. The cryo-buffer is designed to vitrify and prevent crystalline ice formation [28]. Cryo-crystallography experiments are affected by two classes of variables: first, the physical and chemical environment of the crystal in the loop, including choice and concentration of cryo-protectant agent, solvent content of the crystal, size of the crystal and surface area/volume ratio of the crystal; second, the

external variables, including the cooling regime (the cryogen used for flash-cooling, the cryogen flow rate and the temperature used during data collection) and the incident beam conditions of dose, dose rate, flux and wavelength.

The most popular cryoprotectant is glycerol at concentrations between 20 % and 25 %. This does not mean that it is the best cryoprotectant; most experiments involve little optimisation of cryoprotocols and glycerol is often just the easiest available solution that will 'do the job'. The optimum scenario, however, is to grow the crystal in a solution that is already cryoprotected.

Due to faster heat transfer, small crystals are generally easier to flash-cool than larger ones. Larger crystals, however, have a greater diffracting power and will yield more useable diffracted photons before reaching their dose limit. There is thus a minimum crystal size that can be expected to give a data set at a specified resolution for a certain unit cell.

4.5.2 Other Post-Crystallization Treatments

Data collection at cryogenic temperatures, which reduce radiation damage and increase crystal lifetime in the X-ray beam, is essential when dealing with crystals of large macromolecular assemblies. However, the development of protocols for cryo-crystallography in large viruses sometimes fails. Even in sufficiently cryoprotected crystals, rapid freezing can introduce lattice disorders that reduce the diffraction limit. This effect, most likely attributable to the unequal spatial contraction of the solvent and the crystal lattice, is more pronounced for larger crystals with high solvent content [29]. Furthermore, to reduce safety hazards, crystals of infectious pathogenic viruses may have to be mounted in sealed capillary tubes for data collection. In these cases, a large number of crystals will have to be examined in order to solve the structure. Alternative post-crystallization treatments may be helpful to reduce the amount of virus crystals needed for structural studies of human pathogens.

Crystal dehydration can lead to remarkable improvements of diffraction quality. The controlled reduction of the solvent content in a crystal decreases the volume of the unit cell, altering in most cases the crystal packing. These rearrangements may reinforce existing and create new crystal contacts and can thereby lead to better ordered and better diffracting crystals. A sophisticated example of this application has been found in the structural analysis of the membrane-containing bacteriophage PRD1. Crystals of PRD1 were very delicate and the mechanical handling involved in preparing them for data collection rendered them useless for this purpose. In this case a method for growing PRD1 crystals in thin-walled quartz capillary tubes by vapor diffusion were developed, which obviated the need for crystal handling and rendered X-ray data collection practical. These crystals grown in capillaries were further treated by the addition of small amounts of a 20 % (w/v) PEG20K solution injected into the capillary prior data collection [30].

The new developments in techniques for measuring diffraction data directly from crystals in their crystallization plates allows to by-pass a number of problems associated with crystal handling and cryo-protection. By eliminating crystal-handling, plate-based data collection in modern synchrotron beamlines provides a safe platform for the analysis of pathogenic particles, as recently reported in the structure solution of bovine enterovirus 2 [31].

4.5.3 Large Unit Cells and Fragile Crystals

Large unit cell dimensions are inherent to crystals of viruses. Therefore, various crystallographic challenges have to be undertaken when starting with data collection; crystals with large unit cells diffract more weakly because fewer unit cells per crystal volume contribute to scattering. Furthermore, as a consequence of long unit cell axes, the distances between reflections in reciprocal space are small, leading to closely spaced diffraction spots on the detector. Modern undulator beamlines in third generation synchrotrons offer the best instrumentation for successful data collection when working with large-unit cells. Their high beam intensity compensates for the weak diffraction while the small beam divergence and low energy bandwidth decreases spot size, reducing the problem of overlapping reflections. The increased surface area of modern detectors allows crystal to detector distances to be increased. This further reduces the problem of overlapping reflections while increasing the signal to noise ratio as the background radiation is inversely proportional to the square of the detector distance. In addition to the best possible instrumentation, an optimized data collection strategy is needed for a successful experiment. Among the general considerations that should be taken into account, the orientation of unit cell axes is of special importance. If one of the unit cell axes is substantially longer than the others it should be aligned with the spindle axis of crystal rotation to avoid overlapping reflections from neighboring lunes [29].

The strong beam intensity and long exposures necessary to collect high-resolution diffraction data from crystals of large macromolecular complexes causes severe radiation damage and frequently limits the amount of data that can be collected from one crystal, even at cryogenic temperatures. In favorable cases only a few crystals are needed to collect useful data (*i.e.* for frozen crystals of a virus belonging to a highly symmetric space groups). However, for more challenging analyses, a large number of crystals can be required (for example, in the studies on BTV and reovirus cores over 1,000 crystals, mounted in sealed capillaries, had to be examined [9, 32]). Potential problems with non-isomorphism of crystals and crystal quality can be associated with data collection from many crystals. Post crystallization treatments will reduce problems with non-isomorphism and variations in cell dimensions.

The measurement of low-resolution reflections is often overlooked in traditional protein crystallography. For virus crystallography these reflections can be very

important, for instance in using a low-resolution cryo-electron microscopy reconstruction as a phasing model (see Sect. 4.6 and Chaps. 3 and 7), or for visualizing less well ordered structures [33].

4.5.4 Data Processing

The most common data collection method in macromolecular crystallography is the rotation method that involves, as the name indicates, a simple rotation of the crystal around a single axis in small increments (generally between 0.1° and 0.3° for viruses) while the crystal is exposed to X-rays and the reflections are recorded on an area detector (Fig. 4.5).

Once the crystal is centered, the first rotation image is recorded and the raw electronic signal from the detector is corrected by the detector software for detector-specific non-uniformity defects. The corrected raw image includes a header section with relevant experimental parameters, such as detector distance, beam center position, wavelength etc. and is displayed in a graphical display window. The image is then visually inspected for a number of qualitative key parameters of the crystal. Good diffraction implies single, resolved and strong spots, extending far out in the diffraction angle to high resolution.

Indexing is the assignment of a consistent set of unit cell vectors matching the diffraction pattern. In order for indexing to succeed, the position of the incident primary beam on the detector and the crystal-detector distance needs to be known with precision. For crystals with large unit cells, defining the correct starting values for the detector parameters is the most important for quality indexing. Once a unit cell has been selected, its lattice symmetry determines the minimal symmetry of the reciprocal space. Based on lattice symmetry a data collection strategy is determined. The goal of the strategy is to collect at least enough reflections to cover the entire asymmetric unit of the reciprocal space in order to obtain complete data. Once all image frames are collected, the data processing programs must convert the raw pixel intensity of the measured and indexed diffraction spots into a list of properly scaled intensities and their estimated standard errors. The quality of data is normally judged by the agreement of multiple recorded or symmetry related data.

4.6 Phase Determination

The large unit cell dimensions inherent to virus crystals posed also daunting challenges to the determination of phases required to compute the electron density maps from diffraction data using Fourier transformations (Fig. 4.5). However, the first animal virus crystal structures [7, 8] were landmarks confirming the power of density modification, in particular non-crystallographic symmetry averaging, to solve the phase problem, clearing the path towards more accurate and more complex virus structures that has continued till today [15]. In virus X-ray

crystallography the solution to the phase problem usually proceeds in two differentiated steps with first the determination to a certain resolution of some initial phases that are later refined and extended to higher resolution.

This section contains a number of technicalities presenting the approaches used to solve the phase problem in virus crystallography. Readers interested mainly in an overall perspective should only retain the idea of the central importance of the phase problem in diffraction experiments and that some powerful procedures have been developed to solve it in virus crystallography. Interested readers can find more in-depth explanations in the references listed under *Further reading* at the end of this chapter, in particular the last reference.

4.6.1 Initial Phase Determination in Crystals from Icosahedral Virus

To obtain a set of initial phases for an icosahedral virus crystal it is, in general, necessary to determine first the orientations and positions of virus particles in the asymmetric unit of the crystal. Orientations can be determined using the self-rotation function (or even better the locked self-rotation: a self-rotation obeying the restraints imposed by the particle symmetry). Computation of the self-rotation, which requires information only about the structure factors amplitudes, is usually performed in polar coordinates and represented in sections at constants values of the rotation angle κ corresponding to the icosahedral symmetry axes (Fig. 4.5). Once orientations have been determined, information on the position of the center of the viral particles, the translational parameters, can generally be derived from packing considerations. In some cases electron microscopy analysis of thin-sectioned crystals has permitted direct visualization of packing with the determination of the number and position of particles in the unit cell [34]. Often the symmetry of the crystal can constraint the possible orientations and positions of the virus particles by requiring the coincidence of crystal and viral particle symmetries (Fig. 4.4c). For example, when three orthogonal crystal twofold axes superimpose with three twofold axes from the icosahedral particle only two alternative orientations are possible. In this particular case the position of the virus particle is also fixed by the packing that places the center of the icosahedra at the intersection of the three crystal axis. Coincidence of crystal and icosahedral axis reduces the multiplicity of the non-crystallographic symmetry that is 60 when a complete icosahedral particle is found in the crystal asymmetric unit. Fivefold axes are not among the possible crystal symmetries and consequently the fivefold axes from icosahedral particles will remain always as local, non-crystallographic, symmetries.

When orientations and positions of the viral particles in the unit cell are known, initial phases can be obtained using either isomorphous replacement (with heavy atom derivatives) or MR. Isomorphous replacement, historically the first method capable of solving the phase problem in protein crystallography, requires the

preparation of heavy atoms derivatives, which are crystals isomorphous to the ones from the native sample (ideally with the same packing and the same molecular structures), but containing also a few extra heavy atoms (with high atomic numbers) [35]. Diffraction differences between native and isomorphous crystals can then provide information about the phases. In turn, MR requires the availability of independent structural information about the sample that could be used to obtain approximate phases or, at least, some constraints on the possible phases. This independent structural information can be very diverse ranging from atomic models or electron density maps of related samples to the presence of non-crystallographic symmetries. Anomalous diffraction, an alternative phasing methodology used frequently in protein crystallography, has essentially never been applied in virus crystallography mainly due to the experimental difficulties of collecting the weak anomalous signal with enough accuracy from virus crystals.

Preparation of isomorphous heavy atoms derivatives is similar to all protein crystals including virus crystals, but because there are 60 equivalent units in the icosahedral particle, at least 60 positions on the particle are expected to be modified, which results in a large number of vectors (at least 60×59) between heavy atoms within one particle. Information about the non-crystallographic symmetry is taken into account both during the heavy-atoms search, which can be restricted to an asymmetric unit of the icosahedron, and also during the refinement of the heavy-atom parameters [36]. Heavy atom derivatives was the method used to solve the first icosahedral viruses structures such as TBSV, SBMV, STNV or HRV14 [1–3]. However, for about the last 20 years the vast majority of new virus crystal structures have been solved without heavy atoms derivates. Instead a diversity of ingenious molecular replacement approaches in combination with the presence of non-crystallographic symmetry has revolutionized virus crystallography, allowing the explosive rise in the number of crystal structures from viruses that have been determined.

When the atomic coordinates of a structurally similar virus is available, classical MR techniques can be employed to place the searching model in the asymmetric unit of the cell of the unknown structure. In some cases it has been reported that summing two or more homologous structures can provide improved searching models [23]. The calculated structure factors provide then approximate phases and amplitudes, with correlation coefficients between observed and calculated amplitudes often used as an estimation of the quality of phases at a given resolution. Starting at this stage, very accurate phases for the unknown structure can be derived reaching near-atomic resolutions (as described below in Sect. 4.6.2) by density modification (Fig. 4.7), even in the presence of the minimal fivefold non-crystallographic symmetry and low similarity [37]. Improved searching models have also been obtained in conjunction with the information provided by three dimensional cryo-electron microscopy (cryo-EM), where the coordinates of a starting model built by similarity have been trimmed and adjusted to fit the cryo-EM map [38] (see also Chaps. 3 and 7). True quasi-atomic models, obtained by combining the X-ray coordinates of one of the capsid components with a cryo-EM reconstruction of the complete capsid, have provided successful starting models in challenging studies.

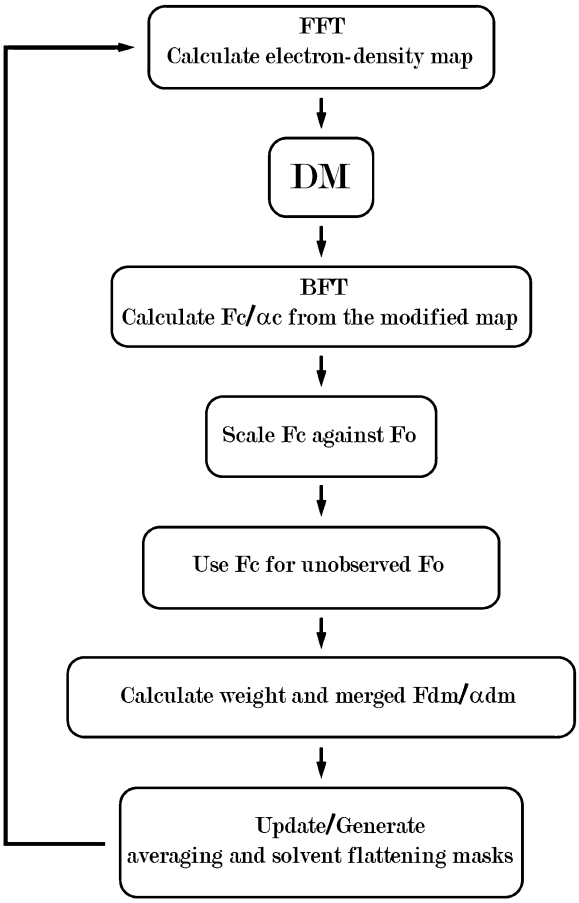


Fig. 4.7 Flow chart of the density modification iterative procedure. Electron density maps, obtained with a Fourier transformation (FFT step), are modified according to the density modification constraints available (DM step), particularly non-crystallographic symmetry averaging and solvent flattening in virus crystallography. The back Fourier transform is then applied (BFT step) to obtain the corresponding structure factors amplitudes and phases (F_c and α_c, respectively). After the experimental and calculated amplitudes are scaled together the calculated structure factors are used to fill the unmeasured reflections. Finally, new amplitudes and phases, ready to provide a new electron density map, are obtained combining and weighting reflections mainly according to differences between experimental and calculated amplitudes. At this point new solvent flattening and averaging masks, defining the volumes where these density constraints can be applied in the next DM cycle, calculated. The process can be repeated iteratively until reaching convergence

After the initial MR solution, the quasi-atomic model is generally improved by rigid body refinement of the capsid components against the experimental diffraction data from the crystal. Using cryo-EM maps of the BTV virion, a quasi-atomic model was built for the outer shell formed by protein VP7, which had been solved by X-ray crystallography [9]. The quasi-atomic model was then placed in the unit cell of BTV

crystals providing initial phases, which were then refined and extended to higher resolution (see below) allowing the accurate structure determination not only of the outer VP7 shell, but also of the inner shell formed by 120 subunits of VP2. Similarly, initial phases were obtained by MR for the structure of the membrane-containing bacteriophage PRD1 using a quasi-atomic model of the viral capsid derived from fitting a cryo-EM reconstruction [13]. Position and orientation of the initial model were refined against the experimental diffraction data of PRD1 crystals, which was followed by rigid body refinement of the 240 P3 trimers included in the model. The PRD1 crystal solved at about 4 Å resolution, allowed to determine the conformation variants adopted by P3 and also the structures of proteins P16 and P31, forming the pentameric base of the vertex spike, and of P30. Initial phases for the crystal structure of human adenovirus solved at 3.5 Å resolution, were also obtained by MR using a quasi-atomic model built by fitting the known hexon and penton base structures into the cryo-EM density of the virus [15]. The 150-megadalton adenovirus capsid with nearly one million residues in the crystal asymmetric unit represents, as indicated before, the largest crystal structure ever solved, confirming the power of MR and non-crystallographic symmetry in virus crystallography.

MR has also been sufficient to solve a number of virus crystal structures when using as searching models the electron density of the virus obtained by cryo-EM. The crystal structure of human hepatitis B virus capsid, with a $T = 4$ triangulation icosahedral lattice, was solved using as searching model the 7 Å resolution cryo-EM map [39]. The initial phases to 8 Å resolution were then extended using density modification procedures to 3.3 Å resolution. At this point it became evident that in the electronic density α helices were left handed, showing that the wrong hand had been selected for the EM map used. Luckily in this case one of the icosahedral twofold axis was coincident with a crystal twofold and the phase extension protocol had not been affected except by producing the inverted hand map. Even the virus masks of cryo-EM maps, with constant density inside, have proved capable to provide searching models for good molecular replacement solutions with initial phases that have been then refined and extended to almost atomic resolution [32].

The feasibility of ab initio phase determination has been demonstrated for icosahedral viruses with initial phases calculated using spherical hollow shell models of the viruses [34, 40]. However, this fascinating possibility has seldom been fulfilled in practice because a number of difficulties mainly related with: (i) the sensibility to certain parameters, such as the outer and inner radius of the shell, (ii) the availability of accurate values for intensities of reflections at low resolution (at least two-thirds of the viral radius), or even to (iii) the packing of the virus particles, particularly when positions present a center of symmetry in the unit cell, as for example when an icosahedral twofold axis coincides with a crystallographic even-fold axis.

4.6.2 *Phase Refinement and Extension: The Power of Non-crystallographic Symmetry*

To improve and extend the resolution of the initial phases, obtained at resolutions lower or much lower than atomic resolution, iterative density modification procedures, in particular non-crystallographic averaging and solvent flattening, can be used (Fig. 4.7). The power of phase determination provided by an iteration of density modification is, according to the semi-quantitative expression derived by Arnold and Rossmann [41], directly related to the square root of the non-crystallographic redundancy (N). Therefore, non-crystallographic symmetry averaging with high non-crystallographic multiplicities, as often available in virus crystallography, can be an extremely powerful mechanism to refine and to extend the initial phases. The accuracy of the observed amplitudes and the proportion of data measured also affect the expected phasing power. In fact, high non-crystallographic redundancy can overcome efficiently the lack of completeness of the experimental diffraction data. Reciprocally, low redundancy can be compensated with the high accuracy and completeness of the measured intensities.

Many of the original problems in virus crystallography related with the averaging algorithms had to deal with limited computer memories and are now essentially solved with the standard software of protein crystallography. It was found that filling the missing experimental amplitudes by including calculated values (suitably weighted) from the back-transformed electron density map can be critical to allow the propagation of phase information and accordingly also critical for the success of the density modification procedure. The power of density averaging is greatly reinforced when used together with solvent flattening, setting the solvent and sometimes also the capsid interior to their respective mean electron densities. In fact, for crystals with a large solvent percentage volume, solvent flattening provides a considerable amount of phasing power, in particular for new reflections added at higher resolution. Thus the envelope or mask defining the solvent must allow the maximum volume without truncating the protein shells. Finding and optimizing the envelopes of the volumes within which the different density modification operations are applied is a critically important aspect of the density modification procedures as already mentioned for the ab initio phasing determination. A number of elaborated alternatives have been developed to refine automatically the envelopes that now are routinely used.

Density modification is mostly applied in real space despite many of the concepts are better controlled in the reciprocal space formulation [42]. In particular, the shell thickness for phase extension should be limited by consideration of the spherical viral particle radius and of the unit cell size. Weighting reflections according to the likely error in their phases can hasten the process. Phases for the newly phased reflections should be allowed a few cycles to converge before the unobserved data are added in. Poor agreement of the strong low-resolution terms may inhibit phase extension. It may therefore be necessary to impose low-resolution cut-offs on the observed data if, for instance, reflections close to the

back-stop shadow have not been measured properly. Termination errors can be reduced by extending the reflections a little beyond the resolution limit of the data, thus allowing the calculated data to contribute to the proper phasing of the observed reflections at the highest resolution. The process of phase refinement and extension is normally monitored by reference to averaging agreement R factors and more reliably to the correlation coefficients between the observed and the calculated structure factors analyzed in terms of resolution.

4.7 Map Interpretation

The final step in the structure determination is the building of a chemically reasonable model that fits the electron density maps generated in the previous steps (Fig. 4.5). Refinement allows completing and optimizing the model maximizing the agreement with all the experimental information available.

4.7.1 Model Building

The accuracy of the phases obtained by the averaging procedure results in high quality electron density maps where the amino acids sequence is often recognized even despite the relatively low resolution (in some cases lower than 3.5 Å) of many of the virus crystal structures determined. In these maps a model is normally constructed for a single icosahedral subunit using standard protein crystallography tools (Fig. 4.5). For viruses with triangulation numbers higher than one the icosahedral subunit contains multiple structurally different copies of a given protein, but once a model for the first protein copy becomes available it can be adapted to provide an initial fit for the electron density corresponding to the remaining copies. Minimaps, skew maps and other map interpretation techniques, widely used in virus crystallography and referred in the literature till the early 1990s, have now been abandoned by the possibilities offered by powerful graphic systems and interactive software such as COOT [43]. For some virus structures solved at low resolution, like PRD1 at about 4 Å resolution [13], averaged difference Fourier maps corresponding to selenium methionine derivatives were calculated using the refined final phases of the native virus. Peaks in these maps, corresponding do the different number of electrons between the selenium and the sulfur atoms (from the selenium and the normal methionines, respectively), flagged the positions of methionines facilitating and providing controls to the sequence assignment during model building.

4.7.2 *Model Refinement*

The coordinates and atomic temperature factors can be refined to minimize the differences between the calculated and observed structure factors while maintaining reasonable stereochemistry. The number of refined parameters can be sharply reduced by applying strict non-crystallographic symmetry constraints resulting in a much higher data-to-parameters ratio, though constrained refinement is only available in a few refinement packages such as CNS and XPLOR [44] or, recently, REFMAC [45]. Since phases derived from the averaging process can be very accurate, these phases can also be used as experimental observations. In fact, due to the quality of the density maps obtained by averaging the aim of refinement is, in general, not to improve the maps, but simply to improve the atomic model. The quality of the averaged maps together with the large computational requirements of virus refinement in reciprocal space explains the successful application of real space refinement to a number of virus structures [3]. From the atomic models a solvent mask is now routinely constructed by most refinement programs, which has a major contribution to the diffraction data at low resolution data (below 8 Å resolution). In cases where high resolution and good data completeness is available the strict non-crystallographic symmetry constraints can be relaxed to analyze deviations from perfect symmetry among the subunits found in the crystal asymmetric unit. This can be particularly important for some external parts of the virus shells involved in crystal contacts [46].

4.8 Understanding Viruses: Some Major Contributions of X-ray Crystallography

The amount, diversity and complexity of viral capsid structures now available at near atomic resolution and stored in the PDB (Protein Data Bank; http://www.rcsb.org/pdb/) or in specialized structural data bases, such as VIPER (VIrus Particle ExploreR; http://viperdb.scripps.edu/), are the result of the outstanding achievements of structural virology, which now treasures the three dimensional structures of over 350 virus capsids, from more than 35 viral families. It is difficult to exaggerate the importance of X-ray crystallography in this process as until very recently X-ray crystallography was the only methodology, and it is still the dominant technique, allowing reaching high resolution with large systems. However, EM is now confirming its enormous potential by providing near-atomic resolution for a rapidly growing number of viral particles (see Chap. 3).

Many more viral protein structures, about 5,000, are also available and a very high percentage of them, more than 90 %, have been determined by protein X-ray crystallography (PDB sept/2012). These includes from structural proteins not forming icosahedral capsids, such as the proteins of cylindrical viruses or the

different components of tailed phages, to many viral enzymes, such as viral proteinases or RNA and DNA polymerases.

The enormous amount of information contained in the structures of virus capsids and proteins, covering many of the relevant issues in virology and allowing many others types of research, is deeply changing our view of the virus world at an accelerated pace. Many of the structures available correspond to potential targets for antiviral drugs and structural virology has had a major role in the development of pharmaceuticals for clinically important viral diseases such as AIDS (see Chap. 20). Drugs directed against the HIV aspartic proteinase are a paradigmatic example of the relevance of structural biology in the development of viral medications, with the structure determination of thousands of crystals from the proteinase complexed with different compounds. The structures of essentially all the protein types encoded by some viruses, such as picornaviruses, have been determined, which allows evaluating the possibilities and limitations of the structural genomics strategies. In fact, it appears clear from these examples that the structural information available for the individual proteins provides only an initial glimpse of the viral functioning as the structures of a diversity of protein complexes and of the proteins interacting with nucleic acids together with alternative protein conformations are still required.

The near-atomic resolution structure and organization of the capsids of some viruses have enriched the information about the underlying building principles expanding the quasi-equivalence symmetry and icosahedral triangulation concepts towards, for example, the interpretation of departures from these principles (as in Polyomavirus or BTV) or to pseudo triangulation arrangements (as in phage PRD1, adenoviruses and many other viruses). This type of information has also contributed to the detailed modeling of the largest spherical viruses, despite they are not yet amenable to high resolution X-ray crystal studies [47, 48].

Specific contributions of X-ray crystallography for understanding particular aspects of stages in the virus life cycle are extremely numerous. We will not deal with particular examples in this chapter, because a substantial number of those contributions have been described in many other chapters of this book, specifically Chaps. 2, 7, 10, 11, 12, 13, 15, 16, 17, 20, 21 and 22. The reader is referred to those chapters to find relevant studies where high-resolution structures of viral particles or components determined by X-ray crystallography have proved instrumental for our understanding of the molecular biology of viruses.

4.9 Perspectives

At synchrotrons, new X-ray detectors and micro-beams together with “in situ” diffraction are leading to a step change that is central to virus X-ray crystallography. Even beyond this, ideas are being developed to grow viral crystals in infected cells, bypassing the requirements for purification and crystallization [49]. Such tiny crystals (typically sub-micrometer) might be amenable to diffraction with advanced

synchrotrons or X-ray free-electron lasers. In fact, the major increases in X-ray brilliance, compared with the most brilliant third-generation SR sources, anticipated from the coming generation of X-ray lasers indicate that the sample size can be reduced to just a single molecule offering the possibility of single virus particles analysis. The oversampling of the continuous scattering originated by the diffraction of a non-periodic sample allows an alternative approach to solve the "phase problem" [50]. In these experiments a sufficiently brief (in the femtoseconds range) and intense X-ray pulse should provide sufficient photons for a useful pattern before radiation damage commences. There are now clear experimental evidences that this single molecule X-ray "diffract and destroy" analysis is feasible and can be applied in particular to virus, though at present only very low resolution has been achieved [51].

Since the first virus structures were determined at almost atomic resolution, about 35 years ago, virus X-ray crystallography has continued providing an always deeper and wider understanding of the virus world with a major impact in the innovation of the X-ray methodology. However, many of the questions remaining in virus structure and functioning will require an even bolder use of X-ray studies in combination with new experimental approaches, such as high-resolution electron microscopies and single-molecule techniques (see also Chaps. 3, 7, 8 and 9).

Acknowledgements We would like to dedicate this chapter to Prof. M.G. Rossmann for his pioneering and continuous ground-breaking results in structural virology. This work was supported by grants from the *Ministerio de Economia y Competitividad* to N.V. (BIO2011-24333) and to I.F. (BFU2009-09268).

References and Further Reading

1. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G (1978) Tomato bushy stunt virus at 2.9 A resolution. Nature 276:368–373
2. Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AG, Rayment I, Rossmann MG, Suck D, Tsukihara T (1980) Structure of southern bean mosaic virus at 2.8 A resolution. Nature 286:33–39
3. Liljas L, Unge T, Jones TA, Fridborg K, Lövgren S, Skoglund U, Strandberg B (1982) Structure of satellite tobacco necrosis virus at 3.0 A resolution. J Mol Biol 159:93–108
4. Verdaguer N, Blaas D, Fita I (2000) Structure of human rhinovirus serotype 2 (HRV2). J Mol Biol 300:1179–1194
5. Verdaguer N, Fita I, Reithmayer M, Moser R, Blaas D (2004) X-ray structure of a minor group human rhinovirus bound to a fragment of its cellular receptor protein. Nat Struct Mol Biol 11:429–434
6. Garriga D, Pickl-Herk A, Luque D, Wruss J, Castón JR, Blaas D, Verdaguer N (2012) Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. PLoS Pathog 8:e1002473
7. Rossmann MG, Arnold E, Erickson JW et al (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317:145–153
8. Hogle JM, Chow M, Filman DJ (1985) Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358–1365

9. Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Ziéntara S, Mertens PP, Stuart DI (1998) The atomic structure of the bluetongue virus core. Nature 395:470–478
10. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289:366–373
11. Roberts MM, White JL, Grütter MG, Burnett RM (1986) Three-dimensional structure of the adenovirus major coat protein hexon. Science 232:1148–1151
12. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648–659
13. Abrescia NG, Cockburn JJ, Grimes JM et al (2004) Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432:68–74
14. Cockburn JJ, Abrescia NG, Grimes JM et al (2004) Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature 2004(432):122–125
15. Reddy VS, Natchiar SK, Stewart PL, Nemerow GR (2010) Crystal structure of human adenovirus at 3.5 A resolution. Science 329:1071–1075
16. Stehle T, Casasnovas JM (2009) Specificity switching in virus-receptor complexes. Curr Opin Struct Biol 19:181–188
17. Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) The structural biology of HIV assembly. Curr Opin Struct Biol 18:203–217
18. Harrison SC (2005) Mechanism of membrane fusion by viral envelope proteins. Adv Virus Res 64:231–261
19. Mukhopadhyay S, Kuhn RJ, Rossmann MG (2005) A structural perspective of the flavivirus life cycle. Nat Rev Microbiol 3:13–22
20. Wang X, Peng W, Ren J et al (2012) A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. Nat Struct Mol Biol 19:424–429
21. Burzlaff H, Gruber B, Wolf PM et al (2002) Crystal lattices. In: Hahn T (ed) International tables for crystallography, vol A. Kluwer Academic Publishers, Dordrecht
22. Fry E, Grimes J, Stuart DI (1999) Virus crystallography. Mol Biotechnol 12:13–23
23. Fry E, Logan D, Stuart D (1996) Virus crystallography. In: Jones C, Mulloy B, Sanderson M (eds) Methods in molecular biology, vol 56: crystallographic methods and protocols. Humana Press, Totowa
24. Vijayachandran LS, Viola C, Garzoni F et al (2011) Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. J Struct Biol 175:198–208
25. Chayen NE, Helliwell JR, Snell EH (2010) Macromolecular crystallization and crystal perfection. Oxford University Press, Oxford
26. Newman J (2005) Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. Acta Crystallogr D Biol Crystallogr 61:490–493
27. Chayen NE (2005) Protein crystallization: automation robotization and miniaturization. In: Sundstrum M, Norin M, Edwards A (eds) Structural genomics and high throughput structural biology. CRC Press/Taylor & Francis, Abingdon
28. Garman E (2003) 'Cool' crystals: macromolecular cryocrystallography and radiation damage. Curr Opin Struct Biol 13:545–551
29. Mueller M, Jenni S, Ban N (2007) Strategies for crystallization and structure determination of very large macromolecular assemblies. Curr Opin Struct Biol 17:572–579
30. Cockburn JJ, Bamford JK, Grimes JM, Bamford DH, Stuart DI (2003) Crystallization of the membrane-containing bacteriophage PRD1 in quartz capillaries by vapour diffusion. Acta Crystallogr D Biol Crystallogr 59:538–540
31. Axford D, Owen RL, Aishima J et al (2012) In situ macromolecular crystallography using microbeams. Acta Crystallogr D Biol Crystallogr 68:592–600
32. Reinisch KM, Nibert ML, Harrison SC (2000) Structure of the reovirus core at 3.6 Å resolution. Nature 404:960–967
33. Gouet P, Diprose JM, Grimes JM et al (1999) The highly ordered double-stranded RNA Genome of bluetongue virus revealed by crystallography. Cell 97:481–490

34. Taka J, Naitow H, Yoshimura M, Miyazaki N, Nakagawa A, Tsukihara T (2005) *Ab initio* crystal structure determination of spherical viruses that exhibit a centrosymmetric location in the unit cell. Acta Crystallogr D Biol Crystallogr 61:1099–1106
35. Baker TS, Johnson JE (1997) Principles of virus structure determination. In: Chiu W, Burnett RM, Garcea RL (eds) Structural biology of viruses. Oxford University Press, New York
36. Arnold E, Vriend G, Luo M, Griffith JP, Kamer G, Erickson JW, Johnson JE, Rossmann MG (1987) The structure determination of a common cold virus, human rhinovirus 14. Acta Crystallogr A43:346–361
37. Fry E, Acharya R, Stuart D (1993) Methods used in the structure determination of foot-and-mouth disease virus. Acta Crystallogr A49:45–55
38. Speir JA, Munshi S, Wang G, Baker TS, Johnson JE (1995) Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. Structure 3:63–78
39. Wynne SA, Crowther RA, Leslie AGW (1999) The crystal structure of the human hepatitis B virus capsid. Mol Cell 3:771–780
40. Tsao J, Chapman MS, Wu H, Agbandje M, Keller W, Rossmann MG (1992) Structure determination of monoclinic canine parvovirus. Acta Crystallogr B48:75–88
41. Arnold E, Rossmann MG (1985) Effect of errors, redundancy and solvent content in the molecular replacement procedure for the structure determination of biological macromolecules. PNAS 83:5489–5493
42. Rossmann MG (1989) The molecular replacement method. Acta Crystallogr A46:73–82
43. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132
44. Brunger AT, Kuriyan J, Karplus M (1987) Crystallographic R factor refinement by molecular dynamics. Science 235:458–460
45. Murshudov GN, Skubak P, Lebedev AA et al (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D67:355–367
46. Garriga D, Querol-Audí J, Abaitua F et al (2006) The 2.6-Angstrom structure of infectious bursal disease virus-derived T=1 particles reveals new stabilizing elements of the virus capsid. J Virol 80:6895–6905
47. Nandhagopal N, Simpson AA, Gurnon JR et al (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. PNAS 99:14758–14763
48. Zhanga X, Xianga Y, Duniganb DD et al (2011) Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid. PNAS 108:14837–14842
49. Abrescia NGA, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral universe. Annu Rev Biochem 81:795–822
50. Raines KS, Salha S, Sanberg RL, Jiang H, Rodríguez JA, Fahimian B, Kapteyn HC, Du J, Miao J (2010) Three-dimensional structure determination from a single view. Nature 463:214–217
51. Seibert MM, Ekeberg T, Maia FR et al (2011) Single mimivirus particles intercepted and imaged with an X-ray laser. Nature 470:78–81

Further Reading

Blow D (2002) Outline of crystallography for biologists. Oxford University Press, Oxford
Rhodes G (2006) Crystallography made crystal clear, 3rd edn. Academic Press, London
Rupp B (2010) Macromolecular crystallography. Garland Science, Taylor & Francis group, LLC, New York

Chapter 5
Nuclear Magnetic Resonance Spectroscopy to Study Virus Structure

José L. Neira

Abstract Nuclear magnetic resonance (NMR) is a spectroscopic technique based in the absorption of radiofrequency radiation by atomic nuclei in the presence of an external magnetic field. NMR has followed a "bottom-up" approach to solve the structures of isolated domains of viral proteins, including capsid protein subunits. NMR has been instrumental to describe conformational changes in viral proteins and nucleic acids, showing the presence of dynamic equilibria which are thought to be important at different stages of the virus life cycle; in this sense, NMR is also the only technique currently available to describe, in atomic detail, the conformational preferences of natively unfolded viral proteins. NMR has also complemented X-ray crystallography and has been combined with electron microscopy to obtain pseudo-atomic models of entire virus capsids. Finally, the joint use of liquid and solid-state NMR has allowed the identification of conformational changes in intact viral capsids on insertion in host membranes.

Keywords Assembly • Binding • Capsid • Chemical shifts • Conformational changes • Coupling constants • Dynamics • Equilibrium • Flexibility • Hydrogen-exchange • Membrane • Nuclear magnetic resonance (NMR) • Nucleic acid • Protein-protein interactions • Protein-nucleic-acid interactions • Solid-state • Solution structure • Unfolded protein • Virus

J.L. Neira (✉)
Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, 03202 Elche, Alicante, Spain

Instituto de Biocomputación y Física de Sistemas Complejos, 50018 Zaragoza, Spain
e-mail: jlneira@umh.es

M.G. Mateu (ed.), *Structure and Physics of Viruses; An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_5,

Abbreviations

BLV	Bovine leukemia virus
CA	Retrovirus capsid protein
CTD	C-terminal domain of CA
COSY	Correlation spectroscopy
EIAV	Equine infectious anemia virus
EM	Electron microscopy
FID	Free induction decay
HIV-1	Human immunodeficiency virus type 1
hLysRS	Human lysyl-tRNA synthetase
HTLV	Human T-cell leukemia virus
MA	Matrix protein
MMTV	Mouse mammary tumour virus
MoMuLV	Moloney murine leukemia virus
MV	Monkey virus
NC	Nucleocapsid
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NTD	N-terminal domain of CA
PIP2	Phosphatidylinositol-4,5-diphosphate
RSV	Rous sarcoma virus
SL	Stem loop
SIV	Simian immunodeficiency virus
SSNMR	Solid-state NMR
TOCSY	Total correlation spectroscopy
VPr	Accessory viral protein R

5.1 Introduction

Although X-ray crystallography (Chap. 4) has been extensively used for determination of the three-dimensional structures of virus particles and viral proteins at atomic resolution, not all biological macromolecules can lead to crystals, and many biomolecules can behave differently in the crystalline state as they do in solution; in those cases, nuclear magnetic resonance (NMR) spectroscopy has been used to determine their high-resolution structures. In addition, the structures of biomolecular complexes, the monitoring of enzymatic reactions, the description of protein folding reactions, or the exploration of the dynamic properties of biomolecules can be addressed by NMR. Therefore, NMR spectroscopy has emerged as an important complement to X-ray crystallography in structural biology in general, and in structural virology in particular.

In this chapter, I shall describe the foundations of nuclear magnetic resonance (NMR) spectroscopy and some of its applications to study viral proteins and virus particles. This technique has a serious limitation when faced to the study of large proteins or macromolecular assemblies: the signals from those biomolecules get too broad to be clearly observed in a spectrum. Due to this limitation, NMR has dealt with the study of whole virus particles, and large viral proteins, by solving the structure of small protein subunits or isolated protein domains of larger proteins (a "bottom-up" approach). NMR has been applied also to study the conformational dynamics of viral proteins in solution, and there have been attempts to tackle the dynamics of complete viral particles. Only recently, solid-state NMR (SSNMR) techniques have been applied to the structural study of whole, intact virus particles in "quasi-solution" conditions.

5.2 Physical Principles of NMR Spectroscopy

5.2.1 The Basis of NMR Spectroscopy

NMR is one of the few spectroscopies that can be explained from both a classic perspective and by using a quantum mechanics point of view. In-depth explanations of NMR theory are beyond the scope of this chapter, and I shall describe briefly only some basic aspects in the following paragraphs. There are excellent books where the interested reader can find a detailed treatment of both complementary views on the NMR foundations [1–5].

NMR is a physical phenomenon based on the magnetic properties of atomic nuclei. Those nuclei with odd numbers of protons and/or neutrons have a magnetic moment (μ) due to their spinning. From a classical point of view, in the presence of an external static magnetic field (B_0), the μ of the nuclei will orient parallel to (with) or anti-parallel to (against) B_0. This interaction originates a precession of the nuclear dipoles around the external field (Fig. 5.1a), in a manner analogous to a gyroscope spinning in a gravitational field. The angular frequency (ω) of this motion depends on the strength of B_0 and the properties of μ. The Boltzmann distribution dictates that the most stable state (nuclear magnetic dipoles oriented with B) will be the most populated at equilibrium, albeit by only a small fraction, because of the small difference in energy with the less stable state (dipoles oriented against B_0). If an appropriate amount of energy is applied, the nuclear magnetic dipoles will move from their equilibrium position and will become oriented against B_0. The corresponding absorption of energy by the nuclei is measured by NMR.

From a quantum perspective, if we consider the simplest nucleus (that of ^{1}H, with a single proton) in the absence of an external magnetic field B_0, the two possible orientations of the nuclear spin (± ½) have an equivalent energy (they are degenerate). In the presence of B_0, two different quantum levels will appear (Fig. 5.1b). The energy difference (E) between those levels is the dot product of

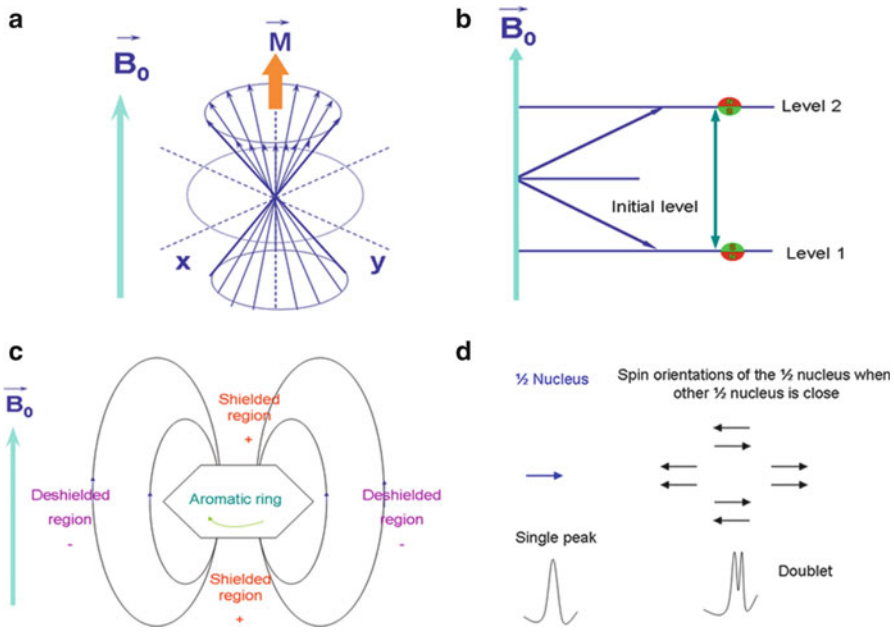


Fig. 5.1 Physical basis of NMR. (**a**) Classical view of NMR. Nuclear spins are precessing around the external magnetic field (B_0, *cyan arrow*); the net magnetization of the system (the sum of all the nuclear magnetic dipoles of the sample) is shown in *orange* (M). (**b**) Quantum view of NMR. The two quantum levels originated as a consequence of the existence of B_0 (*blue arrow*) are shown (for the sake of clarity, it has been assumed that the nuclear spin has two possible orientations only, as in ^{1}H); the orientation of the two spins is shown as *small circles* in each level indicating the polarity of the nuclear magnets. The *double arrow* indicates the jump from one quantum level to another. (**c**) The concept of shielding in aromatic side chains. The *small blue arrows* indicate the direction of the created ring-current magnetic field, for which two lines have also been drawn. The *green line* on the aromatic ring shows the movement of the electron rings. The *red region* is the shielded one, and the *purple zone* is the deshielded region. The direction of the external magnetic field is indicated by a *large cyan arrow*. (**d**) The coupling constant concept for a nucleus with spin ½ (two possible orientations), with a nearby nucleus also with spin ½. From a quantum perspective (see [2–7]), some of those possible orientations are indistinguishable, and they result in two lines (a doublet) in the NMR spectrum instead of the expected single line

B_0 and μ, and it matches the energy within the radiofrequency region; the small separation between both quantum levels is responsible for the low sensitivity of the NMR spectroscopy. To induce a transition between those levels it will be necessary to apply electromagnetic radiation, with an energy given by the Planck equation [1–7]. Thus, the technique is called "nuclear", because we are observing nucleus; it is called "magnetic", because we need a magnetic field to have the different non-degenerate quantum levels; and, it is called "resonance" because we need to apply the exact energy (in the radiofrequency region), provided by the Planck equation corresponding to the energetic separation between the split levels.

In general, nuclei with even numbers of protons and neutrons (such as the most abundant isotope of carbon, ${}_6{}^{12}C$) will have zero spin and they will not yield an NMR spectrum (because there are no levels to split in the presence of B_0). Those nuclei with an odd number of either protons or neutrons will have a half-integer spin; and those with odd numbers of both protons and neutrons will have an integer spin (larger than ½). If the spin of the nucleus is larger than ½ there are more than two quantum states in the presence of B_0 (or, in other words, two different ways of spinning in the classical view), and more spectral transitions, which makes those nuclei much less convenient for NMR analyses than ^{1}H and other nuclei with spin ½. Since the total number of protons and neutrons varies from one isotope to another, it is clear that for the same element there will be some NMR-active isotopes. The most interesting nuclei, from a biomolecular point of view, are: ^{1}H (100 % naturally abundant); ^{15}N, ^{13}C (natural abundance of 0.36 % and 1.11 %, respectively; biomacromolecules need to be artificially enriched in ^{15}N and/or ^{13}C if these isotopes are to be used in NMR experiments; see below); and ^{31}P (100 % naturally abundant; especially useful for studies on nucleic acids).

5.2.2 Chemical Shifts

The atomic nuclei in a molecule are not isolated: they interact with their own electrons, those of other atoms, and with other atomic nuclei. Since the electronic shells have their own spin, they will create an internal magnetic field which will be "felt", in different ways, by the nearby nuclei; therefore, each nucleus in a molecule will have a different "view" (the so-called "shielding") of the external magnetic field B_0, depending on its surroundings. Perhaps the clearest example of this shielding is provided by aromatic rings. In the presence of B_0, a planar aromatic ring with its delocalized π electrons "creates" its own magnetic field (Fig. 5.1c). A nucleus above or below the ring plane will be shielded from the external magnetic field, but it will be unshielded if it is located at other positions relative to the ring; thus, the magnitude of the total field "felt" by a nucleus informs on its location with respect to the ring. In the same way, hydrogen-bonding decreases the electronic shell around a nucleus and reduces the shielding effect. Therefore, it is possible to say that each nucleus in a molecule resonates at a slightly different energy (frequency) relative to others, because of its specific magnetic environment (that is, each nucleus needs a slightly different energy to be promoted to its quantum excited state).

It is impractical to measure the absolute frequency of resonance for every nucleus, and a relative frequency is provided instead. A standard small compound is introduced with the NMR sample, and resonance frequencies of every nucleus in the molecule under study are referred to that of the standard: we measure, in a first stage, the differences between the resonance frequency of each nucleus and that of the standard. Since these measurements of frequencies are provided in hertz (Hz) they are B_0-dependent, and then, the spectrometer magnetic field must be provided.

To avoid an instrument-dependent parameter, a unit-less chemical shift (δ) is provided, given in parts per million (ppm). Thus, δ has the same operational meaning than the wavelength (λ) or frequency (υ) in other spectroscopies (see Chap. 6).

5.2.3 Spin-Spin Coupling: Seeing Other Covalently-Bound Nuclei

The chemical shift, δ, at which a nucleus resonates is a measure of nucleus-electron-clouds interactions. Another important parameter in NMR is the *coupling constant, J*, which is a measure of the magnetic interaction between the spins of two or more covalently bound nuclei (and thus, the interaction occurs *through* the electrons of the bond). The magnetic field "felt" by a nucleus depends on whether the covalently bound nucleus, with a μ, is aligned or not with B_0 (Fig. 5.1d). Therefore, the local field "felt" will be either slightly reduced or increased (when compared to the original magnetic field "felt" when only electronic shielding is taken into account), due to the presence of the neighbouring, covalently-bond nuclei. Then, the original absorption line in the NMR spectrum is split into several lines (two for the case of a single nearby nucleus of hydrogen, in the simplest case, Fig. 5.1d); the strength of the splitting is given by *J*. An NMR-active nucleus covalently bound to nuclei with a null μ (such as protons bound to ${}_6{}^{12}C$) will not show a splitting of its NMR spectral lines; conversely, a nuclei with a nuclear spin larger than ½ (such as ^{14}N) will tend to broaden neighbouring nuclei lines in the NMR spectrum, rather than splitting them.

The *J* is measured in Hz; it is independent of B_0, and depends on the nuclei involved and the number of covalent bonds between the nuclei. For instance, the values of *J* for protons range from 0 to about 20 Hz, but the one-bond *J* for the ^{15}N-^{1}H is close to 90 Hz. As *J* depends on the covalently bound nuclei, it provides information on the dihedral angle connecting them (for protons, for example, this is given by a mathematical relationship termed the Karplus equation [1]). This finding has profound implications, since it allows us to obtain structural information: for instance, the *J* between the amide proton of an amino acid residue in a protein and its H_α proton varies, depending on the type of secondary structure element (α-helix or β-sheet) where the residue is located: 3–5 Hz for residues in α-helices and 8–9 Hz for residues in β-sheets. Therefore, *J* allows to obtain direct structural information on the dihedral angles between covalently-bound atoms.

5.2.4 Relaxation Mechanisms

For simplicity, let us consider a nucleus such as ^{1}H with a nuclear spin of ½, and then, two quantum levels in the presence of a magnetic field. Electromagnetic radiation of the appropriate υ (in the radiofrequency region) causes spins to jump from the lower to the higher quantum level. As the lower level (μ aligned with the

external magnetic field, according to the classical view) has a larger population of spins than the higher one, there will be a net absorption of radiofrequency radiation, leading to a signal in the NMR spectrum. If the electromagnetic radiation is intense enough, the number of spins in the lower and higher levels becomes equal, and therefore, there is no net absorption of energy, because there is no room in the higher level. At the quantum view, we say that the signal is saturated: as many photons are emitted as they are absorbed, and no signal is observed in the NMR spectrum.

However, there are always spontaneous transitions allowing the spins in the saturated higher level to return to the equilibrium population (to the lower level). In other spectroscopic techniques, the recovery of the equilibrium population occurs by "crashing" with other molecules, and losing the excess energy by heat release (see Chap. 6). In NMR, a nucleus relaxes by interaction (through the μ) with a nearby nucleus (from either the same or different molecules); the *spin-lattice* relaxation time, T_1, is the time the equilibrium spin population takes to recover by exchanging energy with the surroundings (*i.e.*, the *lattice*). For a nucleus to contribute effectively to the relaxation of another nucleus, it should be close, and it should belong to a molecule(s) which rotate(s) with an angular frequency similar to the magnetic resonance transition frequency of the studied nucleus (the quantum level separation).

There is another characteristic time, the *spin-spin* relaxation time, T_2, which also measures the recovery of the equilibrium population. The time T_2 refers to interactions *between* spins and measures the rate with which spins exchange energy with each other, which do not involve interactions with the environment (*i.e.*, with the *lattice*). The time T_2 depends, like T_1, on the molecular size and the speed of rotation of the molecule (which is related to the molecular tumbling). Furthermore, T_2 is inversely proportional to the NMR resonance linewidths: the more efficient the T_2-relaxation process, the larger the linewidths (as observed in large molecules, or small ones in rigid environments, for example in SSNMR, see Sect. 5.4.4). Both *T*-relaxation times are important, since they provide information about the dynamics of the molecule.

5.2.5 *The Nuclear Overhauser Effect: Seeing Other Nuclei Through Space*

A very important parameter in NMR is the *nuclear Overhauser effect* (NOE), which is due to a through-space interaction between nuclear spins. An NOE is a change of intensity of an NMR resonance corresponding to a particular nucleus, when a nearby nucleus is irradiated. Those nuclei do not have to be covalently bound; it is necessary and sufficient that they are spatially close enough. Thus, determination of NOEs for different nuclei in a macromolecule allows us to determine which nuclei are near each other in space, a crucial step in determining its three-dimensional structure by NMR (see below).

When intense radiation corresponding to the resonance transition frequency of one nucleus is applied to the sample, spin saturation occurs and, therefore, equal populations of spins are found in the lower and higher levels (as described in Sect. 5.2.4). However, that nucleus is not isolated, it can interact with other nuclei nearby (belonging to the same or different molecule). This results in a change in the equilibrium population of spins. Therefore, when we irradiate a nucleus with the proper energy to cause a transition between its levels (and then, to observe its corresponding peak in the NMR spectrum), we shall observe a variation in the intensity of the peaks of nearby nuclei, compared to the intensity observed if the first nucleus had not been excited. Since the interaction between the corresponding magnetic dipoles depends on r^{-6} (r being the distance between them), an NOE is only observed for close neighbours (usually less than 5Å); the NOE short-range limits its applications, but it provides great specificity in the determination of molecular structures by NMR (see below).

5.2.6 *Pulse Techniques and Multi-dimensional NMR Spectroscopy*

Most spectroscopic techniques (such as fluorescence and CD, see Chap. 6) collect the spectra by varying the wavelength of the excitation source. We could do the same in NMR: varying the wavelength of the radiofrequency source would sequentially induce transitions of different atomic nuclei between their quantum levels (Fig. 5.1b); however, modern NMR spectrometers use radiofrequency pulses, due to improved signal-to-noise ratio and versatility. Since most of the tiny magnetic dipoles – but not all, due to the equilibrium Boltzmann distribution – are aligned with the external field (Fig. 5.1a), there is a net magnetization, M, along the direction defined by B_0 (in the classical view), and most of the spins will start precessing along that direction (see Sect. 5.2.1) (Fig. 5.2a, left side). If we apply a short (microsecond) radiofrequency perturbation (a *pulse*) along the x-axis (Fig. 5.2a), B_1, with a frequency matching the separation energy between the quantum levels (or, in other words, in the classical view, the frequency of precession of the magnetization), the spins aligned initially with the z-axis will start moving away from that axis, and will tend to align with the y-axis (and therefore, in the classical view they will start precessing along that axis). When we remove the perturbation (the pulse) and as the time passes, the different spins forming the net magnetization M, will start to separate at their different precession frequencies (due to their diverse δs), returning to the z-axis (Fig. 5.2a, right side). Thus, the magnetization vector will decrease with a T_1-relaxation time constant (see Sect. 5.2.4.) leading to an exponentially decreasing signal termed a *free induction decay* (FID). The FID is usually observed during a time called t_2 (note that this is an *observation* time, different to the relaxation time T_2) (Fig. 5.2a, right side). Each spin (nucleus) makes a different contribution to the FID due to its slightly

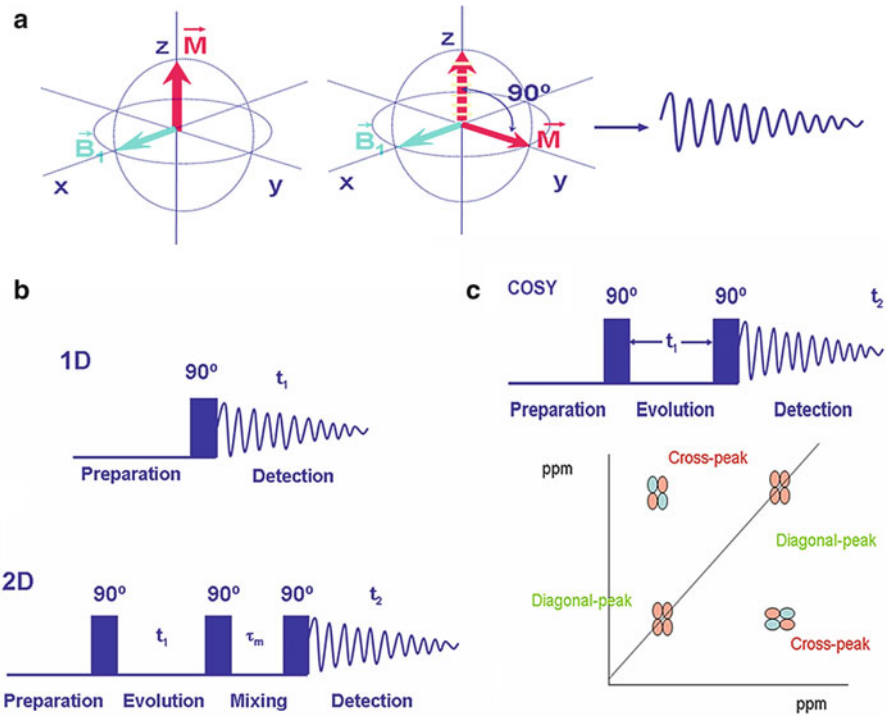


Fig. 5.2 Pulses and n-dimensional NMR. (**a**) How does a pulse work? Initially (*left image*) the system is in equilibrium in the presence of an external magnetic field (which is not indicated); that is, this image shows the state of the system as in Fig. 5.1a. A pulse (B_1, cyan) is applied along the x-axis, and its effect (*right image*) is to tilt the magnetization away from the z-axis. I have represented a pulse which tilts completely the magnetization on the y-axis, but it does not necessarily do so (see, for instance [2–7], for the different lengths of the pulses). From the evolution with time of this tilted magnetization to recover its equilibrium position a FID is obtained (*shown at far right*). (**b**) A series of pulse sequences (pulses are indicated by *rectangles*): *top*, scheme to obtain a 1D-NMR spectrum; *bottom*, scheme to obtain a 2D-NOESY spectrum. The time periods during the experiments are labelled; by convention, the acquisition time during a 1D-NMR spectrum is called t_1 (but it is called t_2 in the 2D-NMR spectra). The numbers on top of the rectangles (the pulses) indicate the tilt caused by the pulse (that is, a 90° pulse tilts the magnetization completely on the y-axis). (**c**) A scheme to obtain a 2D-COSY spectrum. The pulse sequence is depicted on top, where the time t_1 is linearly increased in each experiment. The resulting spectrum showing the diagonal peaks and the cross-peaks is shown at the bottom. The clusters of peaks in a COSY are four-lobed, with alternated signs (positive and negative), due to the effect of the coupling constant among the covalently-bound nuclei

different δ; the individual contributions of the different spins can be separated by means of a Fourier transformation of the time-dependent signal (which relies on the mathematical conclusion that any time-dependent function is equivalent to a frequency-dependent function) [2–7].

With the application of one pulse to the sample, we obtain a one-dimensional NMR (1D-NMR) (Fig. 5.2b, top panel) spectrum, which in graphic terms is, as in

any other spectroscopy, a plot of absorption intensity *versus* δ (or ν). As the molecular size increases, the number of nuclei is larger and, therefore, a large overlapping of δs among the different nuclei will occur when collecting a 1D-NMR spectrum, just because the magnetic environment around some of them will be similar, by chance, even if they are located in different parts of the molecule. Therefore, if we want to identify (and isolate) the δ (or ν) of each particular nucleus, we must try to separate those overlapping frequencies for the nuclei which are found in a similar environment. If a series of pulses are applied to the sample, we can split those overlapping frequencies in two (or more) axes, increasing the spectral resolution; in that way, we obtain a multidimensional spectrum, in which absorption intensity is plotted as a function of two (in 2D-NMR) or more (n-dimensional NMR) δ (or ν) values (Fig. 5.2b, bottom panel). A multidimensional spectrum can be used to monitor the interactions of nuclei through bonds (obtaining *J* values) or through space (obtaining NOEs). The through-bond (or through-space) interactions can be homonuclear (for instance, ^{1}H-^{1}H) or heteronuclear (such as ^{1}H-^{15}N). I shall describe briefly in the next paragraphs two types of homonuclear 2D-NMR experiments; a clear description of some of the most important 2D-NMR experiments used in structure determination can be found in [2–8].

The simplest 2D-NMR spectrum (a homonuclear one called 2D-COSY) is obtained by applying two pulses before observation of the FID during the time t_2 (Fig. 5.2c). The time t_1 that separates the two pulses (Fig. 5.2c, top panel) is varied in successive experiments, and as a result the FID observed during t_2 will change, since the spins will have evolved differently during that small increment in t_1. That is, in the end, we have *k* (the number of times we increase the time t_1) 1D-NMR experiments, where in each of them, a parameter, t_1, is modified. Therefore, the FIDs obtained are a function of two times: t_1 and t_2. A Fourier transformation of this matrix leads to a 2D-NMR spectrum in which resonances are obtained as a function of two frequencies ν (or chemical shifts δ), or "dimensions" (Fig. 5.2c, bottom panel). In the end, we have a contour plot-level (analogous to those used in the weather forecast) where we have the different resonances of protons split in two dimensions (in two axis, as in a Cartesian coordinate map). In the resulting spectrum, each resonance from the 1D produces a correlation with itself, leading the so-called diagonal-peaks (because there is a diagonal through the two opposite vertices of the contour map); these peaks do not yield useful information, since they provide the same information as in the 1D-NMR spectrum. On the other hand, we have off-diagonal cross-peaks, which provide the useful information, since they correlate nuclei (protons, in the 2D-COSY) which are related by (in this case) the covalent-bonds between them (Fig. 5.2c, bottom panel).

On the other hand, a homonuclear 2D-NOESY spectrum consists of the addition of a third pulse, after the last pulse of a COSY, and from which it is separated by a fixed time (mixing time) maintained throughout the experiment (Fig. 5.2b bottom panel); during this time, the spatially close spins are allowed to interact through the NOE (see Sect. 5.2.5). We can also design, for instance, heteronuclear spectra, where instead of observing in both dimensions the same nucleus (the ^{1}H), we observe in the t_2 dimension (the horizontal axis) the ^{1}H and in the t_1 (the vertical

axis) another type of nucleus (such as ^{15}N or ^{13}C). With biomolecules, these heteronuclear experiments (especially, one called HSQC, which shows the correlations between ^{1}H (horizontal axis) and ^{15}N (or ^{13}C) (vertical axis)) are useful to map biomolecule-biomolecule interactions (see below), but it must be kept in mind that due to the low sensitivity of the NMR technique, and the low natural abundance of the latter NMR-active isotopes, the biomolecule must be isotopically enriched in those nuclei.

The key to 2D-NMR spectroscopy is having a time (t_1) which is varied in each of the 1D-NMR spectra acquired (Figs. 5.2b, c), during an observation time, t_2. Therefore, if instead of varying a single time (t_1), we modify two times t among the pulses, we have 3D-NMR. In 3D-NMR we have "cubes" with diagonals between two opposite vertices, and there are off-diagonal, cross-peaks, which contain useful information. To obtain a good resolution, the 3D-NMR experiments require the use of labelled samples. The 3D-NMR spectroscopy is widely used in the assignment (see Sect. 5.3) and structural elucidation of biomolecules. If we modify three times t among the pulses, we have 4D-NMR spectroscopy, and so on for n-NMR spectroscopy. In most of the examples described in the literature, the use of 2D and 3D-NMR spectra with double (^{15}N and ^{13}C) or even triple (^{15}N, ^{13}C and ^{2}H) labelled samples is enough to obtain the structure of a biomolecule.

5.3 Determination of Biomolecular Structures by NMR Spectroscopy

5.3.1 Sample Preparation

Most NMR experiments are carried out in H_2O (to which a 5–10 % of $^{2}H_2O$ is added to monitor the spectrum of ^{2}H during the long acquisition times; see [3–5] for a detailed explanation of such addition, which is beyond the scope of this chapter). One major exception are those experiments aimed at describing the hydrogen-bond scaffolding in a molecule (see Sect. 5.3.2). In some cases, especially when working with peptides or protein fragments, organic solvents are used (such as trifluoroethanol or di-methyl sulfoxide). On the other hand, in proteins associated with membranes or other lipids, the use of detergents, such as sodium dodecyl sulphate (SDS), is required, and in some cases conventional high-resolution NMR methods cannot be used: special techniques as SSNMR can, then, be used to gain structural insight.

As NMR is an insensitive spectroscopic technique, experiments aiming at solving molecular structures tend to be performed at concentrations of about 1 mM (for current spectroscopic configurations). Therefore the biomolecule must be soluble (*i.e.*, monodisperse) at those concentrations, and stable enough during the long acquisition times. The sample is introduced in a tube (made of special, non-paramagnetic substances [3–5]), usually 5-mm-wide, which must be free of

contaminants or other paramagnetic compounds. If the biomolecule is stable, NMR spectra tend to be acquired at low pH, since the hydrogen-exchange of amide protons (on which is based the assignment process, see Sect. 5.3.2) with water is diminished; the buffer must not contain any signal which could obscure any of the resonances of the sample (and then, deuterated buffers are usually employed in homonuclear (^{1}H-detected) experiments). The presence of a high ionic strength can decrease the signal-to-noise ratio, due to the presence of local magnetic fields caused by the ions present in solution. Finally, as the resonance linewidth decreases (and therefore, resolution increases) with temperature, the NMR experiments are acquired at (moderately) elevated temperatures (provided that the biomolecule does not unfold or aggregates).

5.3.2 Structure Determination

The steps in determining the structure of a protein or nucleic acid by NMR follow a straightforward logic. We assume that we know the primary structure of the biomolecule, and we want to obtain the secondary and tertiary structures. There are two ways of assigning a biomolecule: (i) by using the natural abundance of ^{1}H (the homonuclear procedure), which relies on the use, mainly, of 2D-NMR spectroscopy; and (ii) by using labelled samples (the heteronuclear approach), which relies in the use of 3D (or even larger dimensions)-NMR spectroscopy. The difference relies on the type of experiments used to assign the spectra, since the restrictions required to calculate the structure are basically the same.

Although there are many variations to the assignment process (which are beyond the scope of this chapter, and which can be consulted by the interested reader in [8, 9]) either in the homo- or heteronuclear approaches, this process can, in essence, be resumed in a series of steps. First, the multidimensional NMR spectra obtained must be *assigned*. This means that every intensity peak at a defined δ in the spectra must be identified as belonging to one particular nucleus in the molecule.

In the homonuclear approach assignment consists, in a first stage, on the identification of the different types of residues (Fig. 5.3). Using an analogy, during this first step we shall assign a *name* (*e.g.*, Asn, or any other amino acid type) to each different set of signals related with the same amide proton observed in a TOCSY experiment (a more sophisticated type of COSY experiment); but, probably, we shall not be able to put a *surname* to each set of signals assigned to a *name* (amino acid type). In other words, we shall not be able to say whether that assigned residue in the TOCSY spectrum, with the features of an Asn is Asn25, Asn72, Asn80 or Asn100 in the 200-residues long protein we are studying. Then, in a second step, we will assign a *surname* to each set of signals with an already assigned *name*, with the acquisition of a NOESY spectrum, which allows the identification of the connectivity chain through space (see Sect. 5.2.5). Therefore in the homonuclear approach, at the same time we are "naming" and "surnaming" each of the residues of the polypeptide chain, we can conclude, from the pattern of

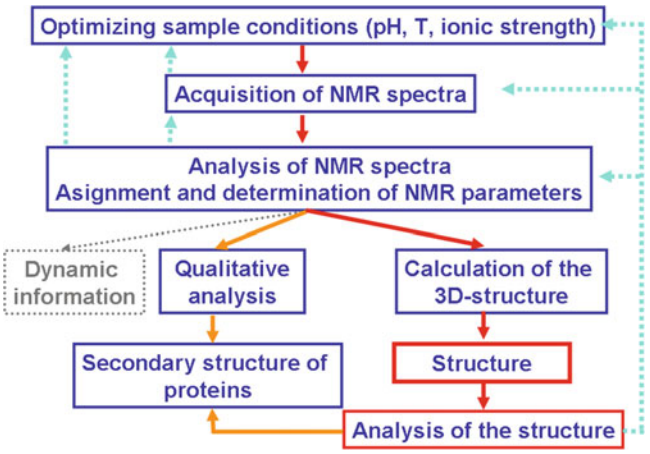


Fig. 5.3 Solving structures by NMR. Flow chart indicating the different steps in the determination of the three-dimensional structure of a protein by NMR

observed NOEs, the type of secondary structure in which the assigned residues are located [2, 5, 7–10] (Fig. 5.3). It is important to note that either the "naming" or "surnaming" steps are mainly based in the presence of amide protons.

In a heteronuclear approach, the assignment step does not involve the use of NOESY spectra [5, 8–10], as a stage required in assignments, since assignment relies on the determination of *J* values for ^{1}H with ^{13}C, ^{1}H with ^{15}N, ^{13}C with ^{15}N, and ^{13}C with ^{13}C nuclei; thus, no ambiguities are introduced, and the assignment process can be relatively easily computerized [9]. In this second approach, NOESYs are used in a late stage, to confirm the through-bond assignment and to provide the data set of restraints (distances between nuclei) for the next step in the structure solving protocol [9] (see below). As it happens with the homonuclear approach, the assignment of the resonances is based (although not exclusively) in the presence of the amide protons.

Once the spectra have been assigned, a number of parameters for the different identified nuclei can be obtained (see Sect. 5.2). The most important are the chemical shifts (δ), coupling constants (*J*), peak intensities, NOEs and, for dynamic studies, the relaxation times (in some cases, even for the heteronuclei). In particular, distances between atomic nuclei (the so-called distance restrictions) can be obtained from the NOEs, and dihedral angles between bound nuclei can be obtained from *J* values. Probably, the most important restriction set to obtain the structure is that provided by the distances, which make the NOESY spectra a key component of the pulse-program tool-kit of every NMR structural determination.

There is another final parameter which is also used in structural determination, because it provides information on the hydrogen-bond scaffolding of the biomolecule (secondary and tertiary structure in proteins, and hydrogen-bonding pattern in nucleic acids). For clarity, I shall focus on proteins in the next lines. In proteins, the amide hydrogens, and other hydrogens bound to nitrogens (as in the side chains of

Lys, Arg, Asn or Gln) are weakly bound due to the hybridization of nitrogen and its electronegativity, as are the hydrogens in the water molecule. Therefore, the N-H bonds (in proteins) and the O-H bonds (in water) can break apart due to the thermal fluctuations, and can immediately be captured either by an oxygen of other water molecule or by a nitrogen of the protein, leading to *hydrogen-exchange* between protein and water. If the protein is dissolved in ^2H_2O instead of H_2O, we should observe an exchange between the labile ^1H of the protein and the ^2H of the solvent. Furthermore, since the resonance (ν) of the ^2H is not the same as that of the ^1H, if we are observing the latter nucleus, we shall not be able to observe the former by NMR; therefore, acquiring spectra at different times after dissolving the biomolecule in ^2H_2O, we shall observe the disappearance of the ^1H, involved in N-H bonds, in those 2D-NMR spectra. However, not all the ^1H will exchange with the same easiness, since those hydrogen-bonded will be more "reluctant" to exchange with the solvent. For structure determination, the information on which protons are more "reluctant" to disappear in the presence of ^2H_2O is extremely worthy, but the speed of exchange also provides information on the stability of the exchanged N-H hydrogen-bonds. Thus, the pattern of exchange of ^1H with ^2H will be hydrogen-bond- and protein-stability dependent, and the number of incorporated ^2H nuclei will allow us to have an estimation of the overall protein stability. A similar approach is used to measure the stability of viruses capsids by using mass spectrometry (see Chap. 6), since the mass of ^2H is roughly twice that of a ^1H, the number of incorporated ^2H nuclei can be easily determined within the precision of current mass spectrometers.

It is important to note that I have described here the structural restrains most frequently used in NMR, but there are other parameters, which are used in partially folded proteins or intrinsically disordered proteins, that are beyond the scope of this chapter. The interested reader can have a wider look in [5, 7, 9].

To produce a structural model of the macromolecule by NMR, we have to find the conformation(s) that is (are) consistent with the experimentally obtained restraints (distances, dihedral angles and hydrogen-bonds) (Fig. 5.3). Usually, these restraints are incorporated into potential energy functions to calculate the molecular dynamics and minimum energy structures [9]. Structures with the fewest violations of the NMR restraints are found by this computing-intensive process. If there are enough restrictions only one structure (or set of very similar structures) can be obtained; with fewer restraints a number of somewhat different structures is found, all of which are consistent with the experimental data. During this computing process, the obtained structures are used to refine the restrictions, and to reassign many of the NOEs observed in the NOESY spectra, which could not be unambiguously determined before (Fig. 5.3, dotted line). The final result is an ensemble of refined, similar structural models that resemble the true minimum energy three-dimensional structure of the macromolecule. Sometimes these models are averaged to provide a representative model of the entire ensemble. Those structures will appear with well-defined regions (where usually there is a large number of structural restrictions *per* amino acid or base) and other zones where the backbone of the protein or the nucleic acid will appear less restricted. These ill-defined regions

may correspond to zones with a high mobility, and as such they are frequently involved in molecular recognition processes; but also, these ill-defined regions might be so due to the absence of local and/or long-range restrictions

5.4 NMR Structures of Viral Macromolecules

In this section, I shall describe several examples of the use of NMR spectroscopy to solve the structure and investigate the dynamics of viral proteins and nucleic acids involved in defining the structure of virus particles (the so-called "structural biomolecules"). A number of proteins from different viruses have been studied by NMR. However, preferred targets have included structural proteins and nucleic acid domains from the human immunodeficiency virus type 1 (HIV-1) or other retroviruses, mainly due to their interest as targets for the development of new antiretroviral therapies (see Chap. 20). In this section, I shall summarize a number of structural studies on retroviral proteins and nucleic acids as case studies to show what NMR can reveal about virus structure. In addition, a few examples on proteins from other viruses will be used to illustrate some other applications of NMR in structural virology.

5.4.1 NMR Structure of Viral Proteins

The main protein components of retrovirus particles are synthesized as three polyproteins that produce the internal structure of the virion (Gag), the viral enzymes (Pol) or the glycoproteins of the virion envelope (Env). The Gag polyprotein is required and sufficient for virus particle assembly and budding. The Gag proteins capture the viral RNA and assemble either in the cytosol (B- and D-type retroviruses) or at the cell membrane (C-type, human T-cell leukaemia virus (HTLV), bovine leukemia virus (BLV), and lentiviruses) to form a RNA-containing spherical capsid in the immature virion. During virion maturation, processing of the HIV-1 Gag polyprotein by the viral protease generates the matrix (MA), the (mature) capsid (CA), the nucleocapsid (NC) and the p6 proteins.

The Capsid Protein of Retroviruses

In HIV-1, the CA domain mediates Gag-Gag interactions that are critical for immature particle formation [11]. HIV-1 CA is also known to interact with a variety of host cell factors, such as cyclophilin A, a host restriction factor and human lysyl-tRNA synthetase (hLysRS) [12]. CA is a two-domain protein, composed of largely helical N-and C-terminal domains (NTD and CTD, respectively) separated by a flexible linker [13, 14]. Upon proteolytic processing during virion maturation, CA

is released from the Gag precursor, and rearranges to form a conical capsid that surrounds the viral genome and associated proteins [13, 15]. The capsid is later disassembled after virus entry into the host cell, to allow uncoating of the viral genome. Mutations that impair CA assembly or alter its stability lead to dramatic reductions in viral infectivity. Assembly and stability of CA is dependent on inter- and intra-molecular interactions that involve the NTD and/or CTD of CA [14, 16]. The NTD (CA residues 1-146), which is a monomer in solution, forms hexamers in the virion capsid; the CTD (CA residues 146-231), which is dimeric in solution, links adjacent hexamers [14]. The atomic structure of CA of HIV-1 has been studied by both X-ray crystallography and NMR using a "bottom-up" approach, based on solving first the structures of the isolated domains. Recently, the NMR structure of the full-length CA has been solved by NMR, confirming the previous results with the isolated domains.

The structure of the HIV-1 NTD. The structures of the HIV-1 NTD determined by NMR and X-ray crystallography ([17] and references therein) are in good agreement. NTD consists of seven α-helices, two β-hairpins, and an exposed, partially ordered loop (Fig. 5.4a). The domain is shaped like an arrowhead; the β-hairpins and loop project from the trailing edge of the arrowhead and the carboxyl-terminal helix projects from the tip. This structure is different to those of other RNA virus coat protein structures [17]. At its N terminus there is a terminal proline, which forms a salt-bridge with a conserved, buried aspartate residue (Asp51). This salt-bridge cannot exist in the Gag precursor polyprotein, and it has been proposed that the β-hairpin itself is important for triggering condensation of the core particle. The binding site for cyclophilin A is located on the exposed loop and encompasses the Pro90, which in the free monomeric domain in solution adopts both *cis* (absent in the X-ray structures) and *trans* conformations due to the loop flexibility. In the X-ray structure of the complex with cyclophilin A, Pro90 adopts a *trans* conformation [17].

The structure of the HIV-1 CTD. The structure of the HIV-1 CTD has been also determined by NMR and X-ray crystallography. CTD is composed of a short 3_{10}-helix, a highly conserved sheet-turn-helix element called the major homology region and four helices (8–11, in the numbering of the intact CA protein) (Fig. 5.4b) which are connected by short loops or turn-like structures. The dimerization interface of CTD is formed by the mutual docking of α-helix 9 from each monomer (with the side chains of each tryptophan (Trp184) deeply buried in the dimer interface), and hydrophobic interactions between the 3_{10}-helix of one subunit and α-helix 9 of the other. Although there are small variations in the crystal and solution structures of dimeric CTD (see next paragraph), the basic lattice-stabilizing dimer interactions in the mature capsid of the virus are the same as in the solution dimer interface.

Although the X-ray and NMR structures of active dimeric CTD reveals basically the same structure within each monomer (the NMR structure was solved in 2009, and the X-ray one in 1996), there are small variations in the structure of α-helix 9 ([14] and references therein). These variations occur at tertiary level (packing of α-helix 9 against other helices within the same monomer), and at quaternary level (the angle formed by the α-helix 9 across the dimer dyad). Furthermore, in solution

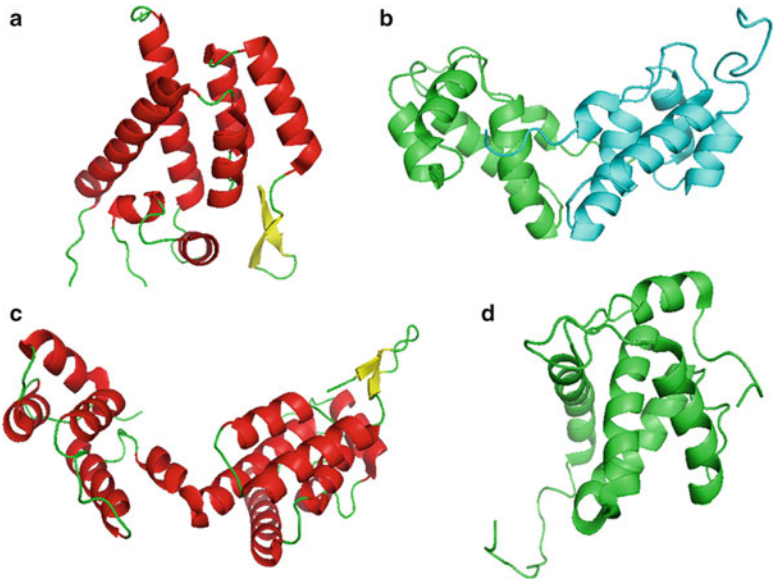


Fig. 5.4 Structures of some HIV-1 proteins solved by NMR. (**a**) Isolated NTD monomer (PDB number 1GWP) [17, 18]. (**b**) Isolated CTD dimer (*monomers are shown in green and cyan*) (PDB 2KOD) [19]. (**c**) Full-length monomeric CA mutant (PDN number 2LF4) [20]; the NTD and CTD are respectively on the *right* or *left side* of the figure. (**d**) The un-myrystilated MA (PDB number 2H3F) [21]. Images in this and following figures were produced with PyMol [22]. In panels (**a**) and (**c**) helices are shown in *red*, β-sheets in *yellow* and loops in *green*

CTD appears to have two conformers, whose population is determined by the protonation of Glu175, which is involved in the dimerization interface [23]. In addition, structurally different dimers have been characterized: namely the one described above, and a domain-swapped arrangement involving mainly the MHR [24]. Somewhat different structures of mutant, monomeric species of CTD have been also determined by NMR. The above and other results, taken together, reveal a high structural plasticity of the CTD that may be biologically important for capsid assembly, maturation and function [14].

Structure of the intact CA proteins of some retroviruses. The structure of the full-length HTLV type I CA and Rous sarcoma virus (RSV) CA have been determined by NMR [25, 26]. As for HIV-1 CA, the proteins contain two independent domains, connected by a flexible linker. Some minor structural differences between the CTDs may account for the fact that HIV-1 CA is dimeric, whereas HTLV-I CA and RSV CA are monomeric. NMR studies of these proteins show also that the two domains are structurally independently, with no inter-domain NOEs; furthermore, the dynamics of the whole domains (as measured by T_1-and T_2- relaxation times) suggest that both regions are also dynamically independent.

The structure of full-length monomeric HIV-1 CA (with an inactive dimerization interface) has also been solved by NMR (Fig. 5.4c). Although the overall

structures of both domains are similar to those reported for the isolated regions, the CTD in whole CA shows differences in the dimerization helix [20]. In the crystal and NMR structures of dimeric CA, the dimerization helix (helix 9), is longer and shows a bend, whereas in the monomeric mutant the α-helix 9 is shorter and straight; this observation supports previous results that suggest substantial conformational rearrangements during CA dimerization ([14] and references therein). The NTD and CTD are connected by a five-residue linker, formed by residues 145–149 (Fig. 5.4c). As observed for other CA structures, the linker is highly flexible, as evidenced by the random coil δs for the central three residues, and the T_1-and T_2-relaxation data. This high interdomain flexibility may be important for facilitating the docking of other Gag domains during immature capsid assembly [27] and allowing the intersubunit structural variations observed in mature retroviral capsids [14].

The Matrix Protein of Retroviruses

The MA protein has similar roles in all retroviruses, directing membrane localization of the assembling viral particle, and after virus maturation, forming a stable structural shell associated with the inner surface of the mature viral lipid envelope ([14] and references therein).

The three-dimensional structures of the HIV-1, BLV and HTLV type II MA proteins have been determined by NMR ([17] and references therein); there are also X-ray structures of the simian immunodeficiency virus (SIV) and HIV-1 MA [17]. The HIV-1 MA consists of five α-helices, two short 3_{10}-helical stretches and a three stranded mixed β-sheet (Fig. 5.4d shows the NMR structure, where only the helices are observed). Close to the C terminus of the second helix there is a putative membrane binding surface, which could anchor the protein to the membrane surface, in concert with an N-terminal myristoyl group. HIV-1 MA is a monomer in solution but forms trimers in the crystal; the HIV-1 MA trimers may be also observed in lower-resolution EM images and could serve as a fundamental building block for formation of the MA shell within the mature virion. The individual proteins of the trimer are arranged to create a basic surface, which interacts directly with the acidic inner layer of the lipid ([17] and references therein). Comparison of a monomer in the crystallographic trimer with the NMR structure of the monomer in solution shows a displacement of the 3_{10}-helix, suggesting that this structural element could act as a “hinge” in virion assembly and disassembly; this resembles the rearrangements proposed for CA dimerization and, again, underscores the importance of protein flexibility in retrovirus morphogenesis. It should be interesting to explore the dynamics of that helix by measurements of T_1 and T_2-relaxation times by NMR, which may serve to unambiguously characterize MA flexibility.

The Nucleocapsid Protein of Retroviruses

NC is the domain of the Gag polyprotein that directs genome packaging. Except for the spumaviruses, all retroviral NC proteins contain one or two copies of a conserved CCHC motif (Cys-*X*2-Cys-X4-His-X4-Cys; where X = non-Cys/His residue), that are Zn-binding sites (the so-called "zinc knuckle"). In HIV-1, NC does not form direct lattice stabilizing interactions, although it contributes to particle formation by tethering Gag to the RNA genome (see [14] and references therein).

NMR structures are available for synthetic peptides with sequences encompassing the HIV-1 NC (the zinc knuckle domains), and for intact NCs from HIV-1, the Moloney murine leukemia virus (MoMuLV) – which contains a single CCHC "zinc knuckle" -, the mouse mammary tumour virus (MMTV), and the monkey virus (MV) [17, 28]. In the HIV-1 and MoMuLV NC proteins, the CCHC zinc knuckle domains adopt similar folds. The N-terminal residues form a metal-coordinating reverse turn called a "rubredoxin knuckle" due to its similarity to metal-coordinating substructures observed in the iron domain of rubredoxin. Subsequent residues form a loop, which leads to a carboxyl-terminal 3_{10}-helix (see Fig. 5.5c for an example, also showing the interaction of NC with a nucleic acid). The two domains in HIV-1 NC show weak NOE inter-knuckle contacts between two bulky aromatic side chains, but dynamic studies (by measuring the T_1- and T_2-relaxation times) suggest that the two "zinc knuckles" behave independently (as it happens with the NTD and CTD in CA, see above). Thus, NC in solution may explore different conformations with interacting and non-interacting knuckle domains [28]. On the other hand, one of the "zinc knuckles" of MMTV and of MV shows a conformational transition involving an additional C-terminal β-hairpin (albeit with different orientation in both proteins). Therefore, the NMR studies suggest that different structural features of the zinc knuckles in diverse retroviruses may facilitate the packaging of the different genomes.

Intrinsically Unfolded Proteins

Here, I shall briefly describe one of the major contributions of NMR to the field of virus structure: the identification and characterization of proteins which are nearly or completely unfolded. Such proteins cannot be crystallized and are not amenable to X-ray structure determination. Currently, NMR is the only technique providing atomic-level insights into this special class of proteins.

The p6 protein of HIV-1 is a docking site for several viral and cellular binding factors, facilitating virus budding [32], and regulating CA processing and virus core assembly [33]. The peptide is disordered in aqueous solution, but under hydrophobic solution conditions (*i.e.*, in the presence of trifluoroethanol) that may mimic a lipidic environment, p6 adopts a helix-hinge-helix structure: a short α-helix (residues 14–18) is connected by a flexible hinge region to a longer helix (residues

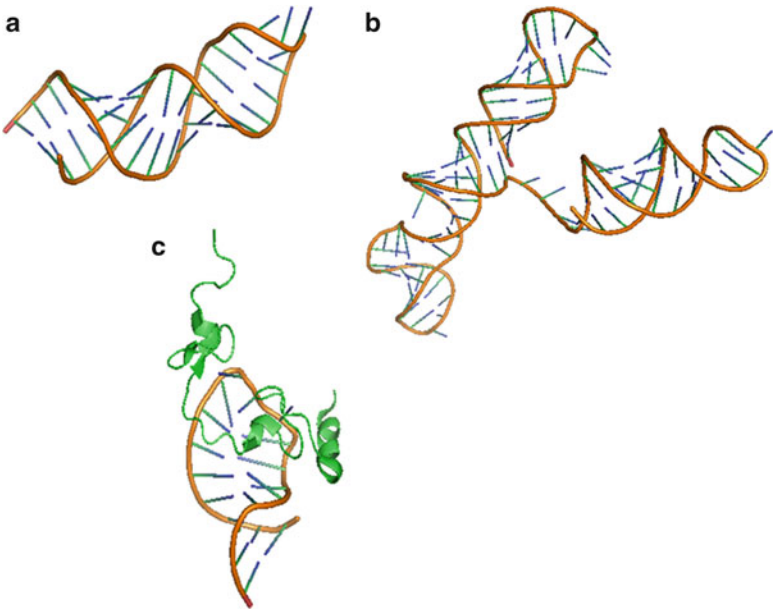


Fig. 5.5 NMR structures of isolated viral RNA or viral RNA-protein complexes. (**a**) Structure of the SL1 of the HIV-1 (PDB number 1N8X) [29]. (**b**) Structure of the MLV Ψ-RNA site (PDB number 1S9S) [30]. (**c**) Structure of the HIV-1 NC bound to SL2 of the RNA (PDB number 1FSU) [31]

33–44). The latter region binds in this conformation to a vacuolar protein sorting factor, and to the HIV-1 accessory viral protein (VPr) [34].

The Gag polyproteins of G-retroviruses contain a conserved p12 domain between MA and CA, which has key roles in virus assembly and nuclear transcription. A MoMuLV fragment, containing the p12 region and the NTD, is unfolded; even the β-hairpin at the N-terminus of the NTD region is unfolded [35]. The p12 region and the hairpin may fold during interaction with other molecules during capsid assembly.

Some non-structural viral proteins contain also substantial intrinsically unfolded regions. For example, the protein W from bacteriophage λ is required for DNA stabilization within the capsid (head) and, attachment of the tail onto the head during virion morphogenesis. The NMR structure of protein W consists of two α-helices and a two-stranded β-sheet arranged around a well packed hydrophobic core, whereas the C-terminal segment, which is essential for DNA stabilization, is unfolded [36]. The 117-residues-long bacteriophage λ FII protein, which is required for the joining of phage head and tail at a late step of phage morphogenesis, has a 3D-structure in solution formed by seven β-strands and a short α-helix. However, two large regions are unstructured: the N terminus (residues 1–24) and a large loop near the middle of the protein (residues 46–62) [37]. The authors suggest that both regions may fold during assembly. As a final example, NMR studies of the

Gn-membrane glycoprotein of hantaviruses show that the zinc finger domain is the sole folded polypeptide patch in the cytoplasmic tail of the protein; this region interacts with the ribonucleoprotein during virus assembly [38].

To sum up, the examples above described, and others found in the literature, suggest that large regions in some structural or non-structural viral proteins are intrinsically unfolded and only become folded upon interaction with other biomolecules. As for other intrinsically unfolded proteins, this lack of order entails an entropic penalty to be paid on binding the ligand; however, this strategy may allow a more promiscuous interaction with different ligands and, in some examples, it could increase the value of the kinetic association rate for complex formation.

5.4.2 NMR Structures of Viral Nucleic Acids

All retroviruses specifically package two copies of their genomes during virus assembly. Packaging is mediated by the Gag proteins, which in the absence of viral RNA, can incorporate cellular RNAs. However, viral RNA contains specific packaging signals (the Ψ-sites), which are preferentially packaged; it seems that genome selection proceeds, albeit not exclusively, *via* the direct binding of NC to those sites. The Ψ-sites are sequestered by intramolecular base pairing in the monomeric RNA, and then, become exposed upon dimerization, allowing high-affinity binding of the RNA dimer to the NC [39, 40]. Probably, a ribonucleoprotein complex formed by a small number of Gag molecules and two copies of the genome is directed towards the plasma membrane assembly sites, where other thousand additional Gag molecules localize and assemble to form the immature virus particle.

Many NMR structural studies on retroviral RNA have focused on the 5′-untranslated region where packaging signals reside. Several studies have shown that the 120 nucleotides upstream of the Gag start codon are required for efficient packaging. This nucleotide segment was proposed to form four stem-loops (the SL1-SL4), with different structures and different affinities for NC [29, 41, 42] (Figs. 5.5a, b). For instance, the NMR structure of the SL2 stem-loop reveals a “platform motif”, where the rings of the sequential U14 and A15 bases are approximately co-planar; this motif is thought to mediate long-range interactions that could stabilize the Ψ-RNA association or facilitate splicing and/or packaging [41]. A similar “platform motif” has been observed in the NMR structure of the 101-nucleotide core encapsidation signal of the MoMuLV, together with an “A-minor K-turn”, where unpaired bases of a bulge pack against the groove of a proximal stem [30].

Recently, the structure of the whole HIV-1 5′-leader RNA (the 712-nucleotide dimer) has been studied by NMR [43]. Residues spanning the Gag start codon (the AUG) form a hairpin in the monomeric leader and base-pair with the bases of the unique-5′ region in the dimer. This association causes dimerization by displacing and exposing a dimer-promoting hairpin, and enhances binding to NC. These and other results taken together suggest a mechanism in which translation,

RNA dimerization, RNA-NC binding and RNA packaging within the assembled capsid are triggered by a common RNA structural switch. This switch would involve the shift of a conformational equilibrium between different conformations in the RNA molecule.

5.4.3 NMR Structures of Viral Protein-Ligand Complexes

Protein-Nucleic Acid Complexes

In RSV, the dynamics of an RNA segment in the free and bound states with a fragment of Gag (comprising CTD, NC and the in-between spacer region (SP)) has been described. Both NC and CTD regions show an independent dynamic behaviour. The contacts with the RNA are distributed along the three regions: one of the zinc fingers in NC seems to make key contacts with the RNA; the SP region experiences a slow conformational equilibrium in the absence of the RNA, which is restrained upon binding; and finally, the first and fourth helices in CTD are also affected by binding to RNA. These results suggest that NC-RNA binding disrupts intramolecular interactions between SP and CTD regions; this disruption enhances the accessibility of the Gag dimer contact surface, promoting intermolecular organization [44].

The structures of several stem-loop sites in the retroviral RNA in complex with the NC of HIV-1 have been determined by NMR [31, 42, 45] (Fig. 5.5c), both zinc knuckles bind to exposed guanine bases of the tetraloop, whereas basic residues at the N terminus of NC pack against the closer "zinc knuckle" being near to the RNA stem; in addition, the N-terminal "zinc knuckle" interacts with the "platform motif" (see above). Some of these interactions are similar to those observed in SL3, but others are substantially different; for instance, the NC interacts with the major groove of SL3, and the orientations of the two zinc domains are different in both complexes [31].

However, not all retrovirus bind with a relative high affinity to the individual SL of RNA; for instance, in MoMuLV, binding of NC only occurs to the three SLs (with a nM affinity) suggesting a high specificity towards a region, which is only present with the intact three loops [42]. The NMR structure of the MoMuLV-NC in complex with the 101-nucleotide core encapsidation segment of the MoMuLV Ψ-site suggests a network of interactions which promote specific binding by the sole "zinc knuckle" of the NC.

Taken together, the structural studies of retroviral RNA-NC complexes reveal that genome recognition and NC-RNA binding differ considerably among the different viruses [39, 43]. Furthermore, these data suggest that the mechanism of genome recognition/packaging involves one or more NC domains of several Gag molecules with multiple Ψ-sites stem-loops [42].

Finally, it is interesting to note that NMR have been used also to detect and map new interactions between viral proteins and nucleic acids without a need to solve the structure of the complex. For instance, MA-DNA interactions have been

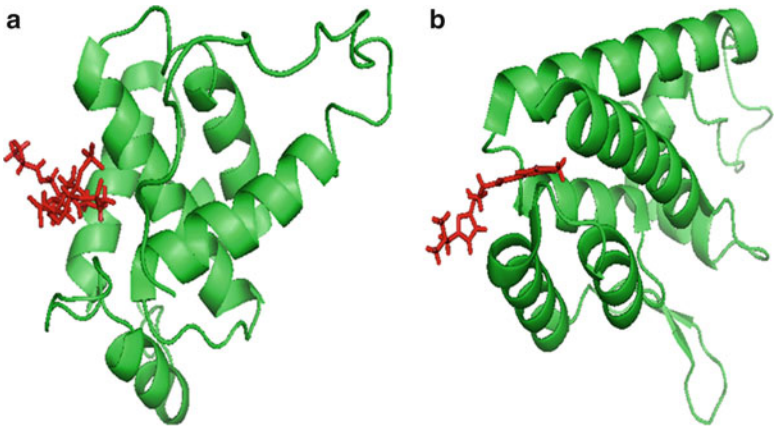


Fig. 5.6 NMR structures of the complexes of small ligands with viral proteins. (**a**) NMR structure of the myristylated HIV-2 MA in complex with the PIP2 lipid (PDB number 2K4I) [48]. The PIP2 lipid moiety is shown in *red sticks*. (**b**) NMR structure, obtained by joint refinement with X-ray data, of the CAP-1 inhibitor in complex with CA of HIV-1 (PDB number 2JPR) [49]. The CAP-1 compound is shown in *red sticks*

detected proved recently to interact with DNA by NMR, but the structure of the complex has not been solved. Instead the DNA-interacting region has been determined by identifying changes in the chemical shifts of the assigned resonances of the protein upon addition of nucleic acid by using heteronuclear HSQCs [46]. This is an example of a more general technique, which allows an easy and very fast detection of protein-ligand interactions in high-throughput screening methods (since it only requires the protein to be labelled with ^{15}N).

Protein-Lipid Complexes

NMR has been used also to study interactions between retroviral MA and lipids. Interactions of MA from equine infectious anaemia virus (EIAV) or HIV-1 with phosphatidylinositol-4,5-diphosphate (PIP2) have been described [47]. The interaction between PIP2 and HIV-1 MA promotes membrane binding and induces conformational changes at the N-terminus of MA (exposing the myristate moiety). The binding region involves the loop between the first and second helices, and the cleft between the second and the fourth helices. In un-myristylated EIAV MA, the binding site of the lipid includes the loop between the second and third helices, although there are conformational changes (as measured by the T_1- and T_2-relaxation times) which may trigger MA trimerization. The structure of myristylated MA of HIV-2 in complex with PIP2 has also been solved [48] (Fig. 5.6a); as it happens with the HIV-1 MA, the myristyl group of HIV-2 MA is partially sequestered within a narrow hydrophobic tunnel formed by side chains of helices 1, 2, 3, and 5, although the myristate of HIV-2 MA is more tightly

sequestered than that of the HIV-1 protein; the addition of PIP2 induces minor changes in the structure of HIV-2 MA. These NMR studies help to understand MA-envelope interactions in retroviruses ([14] and references therein).

Protein-Peptide and Protein-Organic Molecule Complexes

The CA of HIV-1 is being considered as a new therapeutic target (see Chap. 20). To help in understanding how binding to CA of certain peptides and organic molecules may inhibit the assembly of the HIV-1 capsid, the structures of complexes between some of these compounds and CTD of HIV-1 have been solved by NMR ([50] and references therein). Next, I describe two examples.

In the first example, NMR was used to investigate the interactions of peptide CAI, and a stapled derivative (conformationally restricted by a covalent bond between residues that are non-contiguous in the sequence) with a monomeric mutant of CTD. Binding of these peptides to CA results in the inhibition of HIV-1 capsid assembly, and in virion infectivity [50]. The binding interface of CAI with CA was mapped by the changes of chemical shifts of ^{15}N-labelled CA in a HSQC (as it happened with the binding between DNA and MA, see above) ([51, 50] and references therein). The peptides bind in a hydrophobic pocket of CTD delineated by residues of the four helices in the domain, and it adopts a helical conformation where its hydrophobic side chains make extensive contacts with specific hydrophobic patches in CTD. The structural results suggest that binding of these peptides sterically interferes with CTD-NTD interactions in the hexameric rings that form the capsid, and distort the CTD dimerization interface. They also support, once more, the structural plasticity of CTD ([50] and references therein).

The molecule N-(3-chloro-4-methylphenyl)-N´-{2-[({5-[(dimethylamino)-methyl]-2-furyl}-methyl)-sulfanyl]ethyl}-urea) (CAP-1) binds HIV-1 CA and inhibits capsid assembly and virus infectivity. In this example, both X-ray and NMR studies show that CAP-1 binds in a pocket delineated by parts of helices 1, 2, 4 and 7 in NTD [49] (Fig. 5.6b); this region is spatially close to the groove created by α-helices 8 and 9 (the dimerization helix) of CTD. The residue Phe32 of NTD pulls away from its position, creating a large cavity where the aromatic ring of CAP-1 inserts making hydrophobic contacts with other residues in the protein [49]. One general conclusion of these studies is that assembly of the HIV-1 capsid can be inhibited by small molecules (peptides or organic compounds) hydrophobic enough to bind small pockets in CA, to induce local structural distorsions and to hamper CA-CA interactions ([14, 50] and references therein; see also Chap. 20).

This study of CAP-1 in complex with CA also underscores an important aspect of NMR. The joint use of this spectroscopy with other biophysical techniques, such as X-ray crystallography (in the above example) or EM, has led in the last years to promising achievements. For instance, the NMR structure of a viral capsid protein can be combined with the lower-resolution cryo-EM structure of the complete capsid to obtain a pseudo-atomic model. This approach was used in a detailed

structural study of the mature HIV-1 capsid that revealed a small protein-protein interface that had been previously not detected [19]. In another example, the structure of a retroviral RNA packaging element was jointly solved by NMR and cryo-electron tomography [52].

5.4.4 *Structure and Dynamics of Viral Particles by Solid-State NMR Spectroscopy*

The study of intact viral particles and complexes between viral particles and other viral or cellular macromolecules offers a challenge for solution-state NMR and, in some cases, even for X-ray crystallography (see Chap. 4), since they are very large in size, may be not soluble enough, and may prove difficult to crystallize. There have been approaches to tackle the structure of whole virus capsids by NMR, but in all cases, only information about highly dynamic regions of the protein, forming the corresponding capsid, was obtained [53, 54].

In addition to the above approaches using NMR in solution, SSNMR spectroscopy is emerging as a promising technique to study the structure and dynamics of viral particles because it is not limited by solubility, molecular size or lack of long-range order. Major difficulties encountered for obtaining NMR structures of large macromolecules and supramolecular assemblies come from the short T_2-relaxation time, and large linewidth (see Sect. 5.2.4) or, in other words, from the correlation time of the molecule (which is the time used by the molecule in rotating ~57°). The ability of SSNMR to replace molecular motions with radiofrequency irradiations are key to obtain spectra of large molecular assemblies, or aligned molecules within membrane-like environments. The physical details of these methods are out of the scope of this chapter, and the interested reader is referred to recent literature [55–57].

Filamentous bacteriophages are rod-shaped particles encasing a single-stranded circular DNA, within a cylindrical capsid. This capsid consists almost entirely of several thousand copies of the major capsid protein. The virions are not enveloped, but their life cycle involves interactions with the host cell membrane, and the capsid protein also acts as a membrane protein. The best-studied filamentous bacteriophage is fd (class I), which infects *Escherichia coli*. Its capsid is expressed as a procapsid protein, with a 23-residue sequence. *In vivo* the procapsid is inserted into the cell membrane becoming the most abundant membrane protein in infected cells; it is formed by a transmembrane helix connected by a short loop to an amphipatic helix. During virus assembly, the capsid is extruded from the membrane and the DNA is packaged within the capsid. The structure of the capsid protein in filamentous bacteriophage particles was determined by SSNMR [55, 58] (Fig. 5.7a). The first five residues are unfolded and highly mobile. After the Pro6-Ala7 patch the capsid protein is a continuous ideal straight helix between residues 8–38, although it is not a single curved helix: near residue 39 (in the middle of the hydrophobic

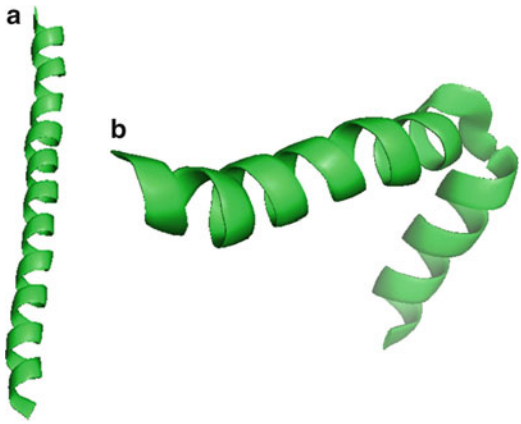


Fig. 5.7 Protein structures solved by SSNMR. (**a**) Structure of a fd capsid protein subunit in intact bacteriophage particles (PDB number 1NH4) [58]. (**b**) Structure of the membrane-bound species of the fd capsid protein (PDB number 1MZT) [59]

helix), there is a kink, causing a change in the helix rotation. Interestingly, the structure of the capsid protein in the virion is not identical to that determined by X-ray fiber diffraction. In the membrane-bound protein, the helix experiences a substantial rearrangement around residue 21 (which was involved in the straight helix in the bacteriophage), that accompanies assembly. The rotation of the trans-membrane helix (residues Gln16-Ala46) around its long axis changes by 16° to obtain the proper alignment for packing in viral particles (Fig. 5.7b) [60].

To allow for a comparison among different bacteriophages, the capsid protein of Pf1 (Class II), which infects *Pseudomonas aeruginosa* has also been studied by SSNMR [55]. The capsid protein of Pf1 is similar to that of fd, with a similar distribution of charged and hydrophobic residues. In contrast to the structure of fd, the six N-terminal residues of Pf1 form a hook; this structure disagrees with the results obtained by X-ray fiber diffraction studies: in fact, multiple chemical shifts were observed for the side-chain and alkyl backbone atoms of Thr5, whereas the rest of the residues had a single chemical shift (suggesting a unique conformation for the 7,300 subunits in the 36 MDa virion [61]. Furthermore, residues around Gln16 (where the hydrophobic helix starts) show a complex dynamic behaviour, with evidence of a bulge (also in disagreement to the X-ray fiber diffraction studies). This behaviour is clearly different to that observed in the capsid protein of fd; instead of a hinge around residue 21, a straight helix was detected.

The dynamic behaviour of the membrane-bound coat protein is also clearly different in Pf1 and fd [55, 60]. For instance, the Pf1 capsid shows a temperature-dependent transition [55, 62]; this transition leads to structural changes in the interfaces between neighbouring capsid subunits, adjusting the hydrophobic interfaces between fairly rigid subunits (as shown by the changes in chemical shifts). Moreover, SSNMR studies of ^{31}P of the DNA backbone suggest that the DNA is packaged differently: in fd the DNA is packaged randomly, while in Pf1 it

is uniformly oriented [54]. Therefore, the SSNMR suggest that both bacteriophages are structurally similar but not identical.

SSNMR has also been used, in conjunction with solution-state NMR, for structural studies of the mature HIV-1 capsid. The capsid structure can be described as a fullerene cone made up from a lattice of CA hexamers containing a number of pentamers at defined positions (see [14] for details). SSNMR studies have shown that both open tubes made only of hexamers and the authentic conical capsid are conformationally homogeneous (even allowing partial assignment of the chemical shifts of several residues in the conical capsid). The structure of CA is not different in both morphologies [16, 23]; furthermore, the structure appears to be similar to that in solution. The linker region connecting the NTD and CTD was, again, found to be very flexible (as shown by the T_1- and T_2-relaxation times) (see Sect. 5.4.1) As already mentioned, this flexibility allows the protein to access multiple conformations essential for assembly of the pleiomorphic retroviral capsid.

5.5 Understanding Viruses: Some Major Contributions of NMR Spectroscopy

In this chapter, I have described how NMR spectroscopy can face the challenging problem of solving the structures of large viral proteins, viral particles and other viral complexes and assemblies by following a "bottom-up" approach. This technique has been used to solve the structure of a number of isolated domains of different viral proteins. The results have generally provided new clues on the relationships between protein structure and viral function, even in cases where the proteins had also been solved by X-ray crystallography. NMR has been essential to reveal the conformational dynamics of several viral proteins, and even of complete viral particles. NMR spectroscopy has shown the presence of alternate conformations, delineating the high intrinsic mobility of viral protein domains, and how such dynamic behaviour contributes to virion morphogenesis. Furthermore, NMR has been revealed as an unmatched technique to show, at atomic level, the natively-unfolded character of some viral proteins. The intrinsic mobility and/or the unfolded character of many (if not all) regions of some viral proteins seem to be landmark features of these viral biomolecules. Probably these features are necessary to allow interactions with other biomolecules (not only during virus assembly, but also at other stages of viral cycle), since a large mobility allow residues to survey a larger number of conformations, and among those, the adequate active structure will be more easily explored. The joint efforts of X-ray crystallography and NMR have allowed to define at a more specific detail the presence of those dynamic equilibria, and even have allowed to identify, and explain some structural ambiguities between both techniques. The combined used of NMR, X-ray crystallography and, in some cases, cryo-EM have allowed a more complete and detailed view of the interactions during assembly. Finally, new advances in SSNMR have contributed to understand the structure of filamentous phage particles and the mechanisms of membrane insertion of these particles.

5.6 Perspectives

A promising future can be envisioned for NMR spectroscopy in structural virology, as exemplified by the techniques and studies described in this chapter, and suggested also by other developments in NMR, which have not been described here because of space limitations. I believe that in the near future, NMR will contribute and/or help in studies dealing with:

i. The structure and, especially, the conformational dynamics of viral proteins.
ii. The structure of viral particles or details of viral proteins in solution through the combined use of cryo-EM (Chap. 3) or X-ray crystallography (Chap. 4) with either solution or solid-state NMR approaches.
iii. The structure and dynamics of viral particles using SSNMR, in conjunction with X-ray fiber or neutron diffraction, for example, to allow a deeper description of the structure of the membrane-bound conformations of viral proteins. Probably, SSNMR could be extended for the description of the first steps of infection (*e.g.*, the formation of fusion complexes).
iv. The structure and dynamics of viral capsids through further advances in solution NMR.
v. The study of structural aspects of a protein within a live cell. The use of in-cell NMR could be applied to viral proteins and assemblies to understand in further structural detail different stages of the life cycle *in vivo*.
vi. The design of new and more efficient drugs to interfere with the activity of any component of the viral machinery (although the first attempts will be probably targeted towards the viral capsid). This will imply the development of new techniques to characterize binding from "the-point-of-view" of the ligand (the so-called *saturation transfer difference* NMR spectroscopy).

Acknowledgements I thank Luis Pérez and Francisco N. Barrera for careful reading of the manuscript, suggestions and new ideas. I apologize to those colleagues whose work was not cited due to lack of space or inadvertent omission. Work in the author's laboratory was supported partially by the Spanish Ministerio de Ciencia e Innovación (MCINN) (CTQ2011-24393 and CSD2008-00005), the FIPSE foundation (2006/0008), and intramural BIFI 2011 projects.

References and Further Reading

1. Atkins PW, de Paula J (2008) Physical chemistry for the life sciences. Oxford University Press, Oxford
2. Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York
3. Derome AE (1987) Modern NMR techniques for chemistry research. Pergamon Press, Oxford
4. Claridge TDW (1999) High-resolution NMR techniques in organic chemistry. Pergamon Press, Oxford
5. Cavanagh J, Fairbrother WJ, Palmer AG III, Skelton N (1996) Protein NMR spectroscopy. Principles and practice, 1st edn. Academic Press, New York
6. Keeler J (2006) Understanding NMR spectroscopy, 2nd edn. Wiley, New York

7. Evans JE (1996) Biomolecular NMR spectroscopy. Oxford University Press, Oxford
8. Bax A, Lerner L (1986) Two-dimensional NMR spectroscopy. Science 232:960–967
9. Reid DG (ed) (1997) Protein NMR techniques. Methods in molecular biology, vol 60. Humana Press, Totowa
10. Dingley AJ, Lorenzen I, Grötzinger J (2008) NMR analysis of viral protein structures. In: Foster GD, Johansen IE, Hong Y, Nagy PD (eds) Plant virology protocols: from viral sequence to protein function. Humana Press, Totowa
11. Adamson CF, Freed EO (2007) Human immunodeficiency virus type 1 assembly, release and maturation. Adv Pharmacol 55:347–387
12. Mascarenhas AP, Musier-Forsyth K (2009) The capsid protein of human immunodeficiency virus: interactions of HIV-1 capsid with host protein factors. FEBS J 276:6118–6127
13. Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) The structural biology of HIV assembly. Curr Opin Struct Biol 18:203–217
14. Ganser-Pornillos BK, Yeager M, Pornillos O (2012) Assembly and architecture of the HIV. Adv Exp Med Biol 726:441–465
15. Adamson CS, Salzwedel K, Freed EO (2009) Virus maturation as a new HIV-1 therapeutic target. Expert Opin Ther Targets 13:895–908
16. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, Stout CD, Sundquist WY, Hill CP, Yeager M (2009) X ray structures of the hexameric building block of the HIV capsid. Cell 137:182–192
17. Turner BG, Summers MF (1999) Structural biology of HIV-1. J Mol Biol 285:1–32
18. Tang C, Ndassa Y, Summers MF (2002) Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. Nat Struct Biol 9:537–543
19. Byeon IJ, Meng X, Jung J, Jung J, Zhao G, Yang R, Ahn J, Shi J, Concel J, Aiken C, Zhang P, Gronenborn AM (2009) Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. Cell 139:178–190
20. Shin R, Zhou YM, Krishna NR (2011) Structure of a monomeric mutant of the HIV-1 capsid. Biochemistry 50:9457–9467
21. Saad J, Miller J, Tai J, Kim A, Granam RH, Summers MF (2006) Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc Natl Acad Sci USA 103:11364–11369
22. DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, San Carlos
23. Byeon IJ, Hou G, Han Y, Suiter CL, Ahn J, Jung J, Byeon CH, Gronenborn AM, Polenova T (2012) Motions on the millisecond time scale and multiple conformations on HIV-1 capsid protein: implications for structural polymorphism of CA assemblies. J Am Chem Soc 134:6455–6466
24. Ivanov D, Tsodikov OV, Kasanov J, Elnberger T, Wagner G, Collins T (2007) Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. Proc Natl Acad Sci USA 104:4353–4358
25. Khorasanizadeh S, Campos-Olivas R, Summers MF (1999) Solution structure of the capsid protein from the human T-cell leukemia virus type-I. J Mol Biol 291:491–505
26. Campos-Olivas R, Newman L, Summers MF (2000) Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. J Mol Biol 296:633–649
27. Jiang J, Ablan SD, Derebail S, Hercick K, Soheilian F, Thomas JA, Tang S, Hewlett I, Nagashima K, Gorelick RJ et al (2011) The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. Virology 421:253–265
28. Klein DJ, Johnson PE, Zollars ES, De Guzman RN, Summers MF (2000) The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the C-terminal zinc knuckle. Biochemistry 39:1604–1612
29. Lawrence DC, Stover CC, Noznitsky J, Wu Z, Summers MF (2003) Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. J Mol Biol 326:529–542

30. D'Souza V, Dey A, Habib D, Summers MF (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. J Mol Biol 337:427–442
31. Amarasinghe GK, De Guzmán RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF (2000) NMR structure of HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Psi-RNA packaging signal. Implications for genome recognition. J Mol Biol 301:491–511
32. Yu XF, Matsuda Z, Yu QC, Lee TH, Essex M (1995) Role of the C terminus Gag protein in human immunodeficiency virus type I virion assembly and maturation. J Gen Virol 76:3171–3179
33. Votteler J, Neumman L, Hahn S, Hahn F, Rauch P, Schmidt K, Studtrucker N, Solbak SM, Fossen T, Henklein P et al (2011) Highly conserved serine residue 40 in HIV-1 p6 regulates capsid processing and virus capsid assembly. Retrovirology 8:11
34. Fossen T, Wray V, Bruns K, Rachmat J, Henklein P, Tessmer U, Maczurek A, Klinger P, Schubert U (2005) Solution structure of the human immunodeficiency virus type 1 p6 protein. J Biol Chem 280:42515–42527
35. Kyere SK, Joseph PR, Summers MF (2008) The p12 domain is unstructured in a murine leukemia virus p12-CA(N) Gag construct. PLoS One 3:e1902
36. Maxwell KL, Yee AA, Booth V, Arrowsmith CH, Gold M, Davidson AR (2001) The solution structure of bacteriophage lambda protein W, a small morphogenetic protein possessing a novel fold. J Mol Biol 308:9–14
37. Maxwell KL, Yee AA, Arrowsmith CH, Gold M, Davidson AR (2002) The solution structure of the bacteriophage head-tail joining protein. J Mol Biol 318:1395–1404
38. Estrada DF, Conner M, Jeor SC, Guzmán RN (2011) The structure of the hantavirus zinc finger domain is conserved and represents the only natively folded region of the Gn cytoplasmic tail. Front Microbiol 2:251–262
39. D'Souza V, Summers MF (2005) How retroviruses select their genomes. Nat Rev Microbiol 3:643–655
40. D'Souza V, Summers MF (2004) Structural basis for packaging the dimeric genome of Moloney murine leukemia virus. Nature 431:586–590
41. Amarasinghe GK, De Guzmán RN, Turner RB, Summers MF (2000) NMR structure of stem-loop SL2 of the HIV-1 Psi-RNA packaging signal reveals a novel A-U-A base-triple platform. J Mol Biol 299:145–156
42. Lu K, Heng X, Summers MF (2011) Structural determinants and mechanism of HIV-1 genome packaging. J Mol Biol 410:609–633
43. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, Divakaruni SS et al (2011) NMR detection of the structures in the HIV-1 5′ leader RNA that regulate genome packaging. Science 334:242–245
44. Taylor GM, Ma L, Vogt VM, Post CB (2010) NMR relaxation studies of an RNA-binding segment of the rous sarcoma virus gag polyprotein in free and bound states: a model for autoinhibition of assembly. Biochemistry 49:4006–4017
45. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. Science 279:384–388
46. Cai M, Huang Y, Craigie R, Clore GM (2010) Structural basis for the association of HIV-1 matrix protein with DNA. PLoS One 5:e15675
47. Chen K, Bachtiar I, Psizcek G, Bouamr F, Carter C, Tjandra N (2008) Solution NMR characterizations of oligomerization and dynamics of equine infectious anemia virus matrix protein and its interaction with PIP2. Biochemistry 47:1928–1937
48. Saad JS, Ablan SD, Ghanam RH, Kim A, Andrews K, Nagashima K, Soheilian F, Freed EO, Summers MF (2008) Structure of the myristylated human immunodeficiency virus type 2 matrix protein and the role of the phosphatidylinositol-(4,5)-biphospahte in membrane targeting. J Mol Biol 382:434–437

49. Kelly BN, Kyere S, Kinde I, Tang C, Howard BR, Robinson H, Sundquist WI, Summers MF, Hill CP (2007) Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. J Mol Biol 373:355–356
50. Neira JL (2009) The capsid protein of human immunodeficiency virus: designing inhibitors of capsid assembly. FEBS J 276:6110–6117
51. Bocanegra R, Nevot M, Doménech R, López I, Abián O, Rodríguez-Huete A, Cavasotto CN, Velázquez-Campoy A, Gómez J, Martínez MÁ et al (2011) Rationally designed interfacial peptides are efficient *in vitro* inhibitors of HIV-1 capsid assembly with antiviral activity. PLoS One 6:e23877
52. Summers MF, Irovalieba RN, Tolbert B, Smalls-Mantey A, Iyalla K, Loeliger K, D'Souza V, Khant H, Schmid MF, Garcia EL et al (2010) Structure of a conserved retroviral RNA packaging element by NMR spectroscopy and cryo-electron tomography. J Mol Biol 404:751–772
53. Vriend G, Hemminga MA, Verduin BJM, De Wit JL, Schaafsma TJ (1981) Segmental mobility involved in protein-RNA interaction in cowpea chlorotic mottle virus. FEBS Lett 134:167–171
54. Szymczyna BR, Gan L, Johnson JE, Williamson JR (2007) Solution NMR studies of the maturation intermediates of a 13 MDa viral capsid. J Am Chem Soc 129:7867–7876
55. Opella SJ, Zeri AC, Park SH (2008) Structure, dynamics, and assembly of filamentous bacteriophages by nuclear magnetic resonance spectroscopy. Annu Rev Phys Chem 59:635–637
56. Sun S, Han Y, Paramasivam M, Yan S, Siglin AE, Williams JC, Byeon IJ, Ahn J, Gronenborn AM, Polenova T (2012) Solid-state NMR spectroscopy of protein complexes. Methods Mol Biol 831:303–331
57. Duer MJ (2004) Introduction to solid state NMR spectroscopy. Blackwell, London
58. Zeri AC, Mesleh MF, Nevzorov AA, Opella SJ (2003) Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy. Proc Natl Acad Sci USA 100:6458–6463
59. Marassi FM, Opella SJ (2003) Simultaneous assignment and structure determination of a membrane protein from NMR orientational restrains. Protein Sci 12:403–411
60. Park SH, Marassi FM, Black D, Opella SJ (2010) Structure and dynamics of the membrane-bound form of Pf1 coat protein: implications of structural rearrangements for virus assembly. Biophys J 99:1465–1474
61. Goldbourt A, Gross B, Day LA, McDermott AE (2007) Filamentous phage studied by magic-angle spinning NMR: resonance assignment and secondary structure of the coat protein in Pf1. J Am Chem Soc 129:2388–2344
62. Goldbourt A, Day LA, McDermott AE (2010) Intersubunit hydrophobic interactions in Pf1 filamentous phage. J Biol Chem 285:37051–37059

Further Reading

For the basis of NMR:

Brey WS (ed) (1996) Magnetic resonance in perspective: highlights of a quarter of century. Academic Press, New York

Homans SW (1992) A dictionary of concepts in NMR. Oxford University Press, Oxford

Neuhaus D, Williamson MP (2000) The nuclear overhauser effect in structural and conformational analysis, 2nd edn. Wiley-VCH, New York

For intrinsically disordered proteins in viruses:

Dyson HJ (2011) Expanding the proteome: disordered and alternatively folded proteins. Q Rev Biophys 44:467–518

Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN (2010) Viral disorder or disordered viruses: do viral proteins posses unique features? Protein Peptide Lett 17:932–951

Also especially recommended for further reading are references [2–9, 57] *listed above*.

Chapter 6
Fluorescence, Circular Dichroism and Mass Spectrometry as Tools to Study Virus Structure

José L. Neira

Abstract Fluorescence and circular dichroism, as analytical spectroscopic techniques, and mass spectrometry as an analytical tool to determine the molecular mass, provide important biophysical approaches in structural virology. Although they do not provide atomic, or near-atomic, details as electron microscopy, X-ray crystallography or nuclear magnetic resonance spectroscopy can do, they do provide important insights into virus particle composition, structure, conformational stability and dynamics, assembly and maturation, and interactions with other viral and cellular biomolecules. They can be used also to investigate the molecular determinants of virus particle structure and properties, and the changes induced in them by external factors. In this chapter, I describe the physical bases of these three techniques, and some examples on how they have helped us to understand virus particle structure and physicochemical properties.

Keywords Assembly • Calorimetry • Capsid • Chemical denaturants • Chirality • Circular dichroism • Conformational changes • Dissociation • Dynamics • Flexibility • Fluorescence • Interactions • Intermediate • Ligand • Mass spectrometry • Protein stability • Refractive index • Structure • Virus

Abbreviations

ANS	1-anilino-8-naphtalene sulfonate
bisANS	bis(4-anilinonaphtalene-5-sulfonic acid)
CD	Circular dichroism

J.L. Neira (✉)
Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, 03202 Elche, Alicante, Spain

Instituto de Biocomputación y Física de Sistemas Complejos, 50018 Zaragoza, Spain
e-mail: jlneira@umh.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_6,

DSC	Differential scanning calorimetry
EM	Electron microscopy
ESI	Electrospray ionization
FHV	Flock house virus
GdmCl	Guanidine hydrochloride
HBV	Hepatitis B virus
HIV	Human immunodeficiency virus
HK97	Hong-Kong 97 virus
HPV	Human papilloma virus
HRV	Human rhinovirus
HSV	Herpes simplex virus
HX	Hydrogen-exchange
m/z	Mass-to-charge
MALDI	Matrix-assisted laser desorption ionization
MS	Mass spectrometry
MVM	Minute virus of mice
NMR	Nuclear magnetic resonance
PKC	Protein kinase C
RYMV	Rice yellow mottle virus
SL	Stem loop
T_m	Thermal denaturation midpoint
TMV	Tobacco mosaic virus
TOF	Time of flight
UV	Ultraviolet

6.1 Introduction

In Chap. 5 I have described nuclear magnetic resonance (NMR) spectroscopy, a powerful structural and analytical technique to study viruses and viral proteins in solution. In this chapter, I shall describe some other analytical techniques in solution, including fluorescence and circular dichroism (CD) spectroscopies and mass spectrometry (MS). These techniques do not provide information at or close to atomic resolution, as electron microscopy (EM) (Chap. 3), X-ray crystallography (Chap. 4) and NMR (Chap. 5) can do; but they allow us to investigate several aspects of virus structure and properties such as stability and conformational dynamics.

In the first part of this chapter, I shall describe fluorescence spectroscopy and CD techniques. Both are faster (experiments can be completed in minutes) and require lower amounts of biomolecules than EM, X-ray crystallography or NMR. Thus, they can be used as a first approach to obtain some information on the structure and

properties of the biomolecule, which can, then, be studied in structural detail by using the latter techniques. Both fluorescence and CD are based in the absorption of light, which induces transits between the electrons of the external shells of the molecules involved. Their uses to study virus structure, stability, assembly, and interactions are described here.

In the second part of this chapter, I shall describe the foundations and some uses of MS in structural virology. This technique is based on changes in the trajectory experienced by ions derived from the target molecule in the presence of electric and magnetic fields. Contrary to fluorescence, CD or NMR, MS does not use the electromagnetic radiation to induce leaps among quantum levels, and then it is not a spectroscopic technique. From the data obtained, MS provides information about the mass of the target biomolecules, even of intact virus particles. Data acquisition is very fast, and the amounts of sample required are even lower than for CD or fluorescence spectroscopies. The use of MS in virus structure, stability, dynamics and assembly, and virus-ligand interactions are also described here.

6.2 Physical Principles of Fluorescence and Circular Dichroism (CD) Spectroscopies

6.2.1 Fluorescence

A detailed account of the general principles of fluorescence can be found in the book by Lakowicz [1], and its applications to biochemistry in the textbook by Albany [2]. I shall briefly revise some basic principles to explain the basis of fluorescence.

The energy of a couple of bound atoms (E(total)) depends on: E(total) = E(translation) + E(rotational) + E(vibrational) + E(electronic) + E(electronic spin orientation) + E (nuclear spin orientation).

The translational energy involves the changes in the position of the molecule; the rotational energy involves the molecular rotation around the gravity center of the molecule; the vibrational energy involves the deformation (compression or stretching) of bonds and angles; the electronic energy involves the movement of electrons within the molecular shell. The last two terms involve the change of the spin of the electron and the change of the nuclear spin (see Chap. 5), respectively. The determination of changes in each of those types of energy leads to different kinds of spectroscopies; the fluorescence and CD techniques are based in electronic transitions.

Each of those types of energies does not distribute continuously between the two atoms, but are "quantized". For instance, the energetic separation between the rungs in the electronic energy ladder is 10^{-17} J; those in the vibrational one are 10^{-20} J; those in the rotational one are 10^{-24} J, and those in the translational one are 10^{-36} J. This energetic distribution leads to two important conclusions. First, inducing an

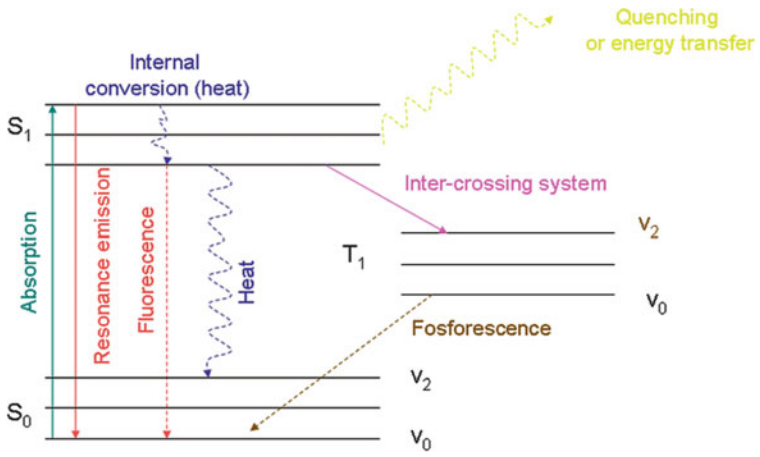


Fig. 6.1 The electronic energy levels and transitions in a molecule. Two electronic levels are shown (S_0 and S_1) with the corresponding vibrational levels (the number of vibrational levels drawn is arbitrary, and they are called $V_0 \ldots V_i$). The absorption of energy is shown by a continuous *cyan line*, and the emission by a *red* one; the internal conversion processes (such as the emission of heat) are shown by *curved dotted lines*. The emission fluorescence is shown by a *dotted red line*. The inter-crossing system, which involves the jumping from an electronic state to another unstable one (T1) with different spin properties, is shown in *pink*, and the fosforescence emission in a *dotted brown line*. The quenching or energy transfer is shown by a *curved dotted yellow line*. The meaning and nomenclature of both electronic terms, S and T, are beyond the scope of this book; the interested reader can have a wider look on the meaning of the electronic states in [1]

electronic transition involves also the induction of rotational, vibrational and translational transitions, because they have smaller energy. And second, for an electronic level there are vibrational ones; for each vibrational level there are rotational ones, and so on.

The Basis of Fluorescence

We deal in fluorescence with photon absorption and electrons located in specific orbitals in a molecule. The absorption of photons induces leaps from a ground electronic level to an excited state. In this excited state, the molecule can return to the ground state following any of these processes (Fig. 6.1):

(a) The molecules dissipate the excess of energy in the surrounding environment, usually, but not exclusively, as heat.
(b) The molecules can release the same amount of absorbed energy by the emission of a photon. This is called resonance.
(c) Part of the absorbed energy is dissipated in the medium as heat. Then, the molecules are in the lower vibrational level of the electronic excited state, and from this rung in the energetic ladder, they release a photon to return to the ground electronic level. The released photon will have a lower energy than the absorbed one, and then, its wavelength will be larger. This process is fluorescence.

(d) A transient passage occurs to an unstable electronic state, from where the molecule can pass to the ground state, by photon emission. This process is phosphorescence.
(e) Some molecules in the excited state can release energy to nearby molecules (by collisional quenching or energy transfer).

It is clear that the fluorescence emission spectrum is the mirror image, shifted towards longer wavelengths, of the absorption spectrum, provided that the vibrational distribution in both ground and excited states is basically the same. However, we cannot conclude that all the absorbing molecules will have fluorescence, because energy can be dissipated within the environment as collisional or thermal energy, or as energy transfer to other molecules. For instance, the heme group does absorb, but it does not show fluorescence, due to transfer of energy from the porphyrin ring to the iron.

Fluorescence of Biomolecules

In proteins, tryptophan, tyrosine and phenylalanine residues are responsible for the UV properties of the molecule, but whereas Trp absorbs more intensely than the two others, Tyr has the largest fluorescence. When both residues are present in a protein, Tyr fluorescence is weaker, due mainly to energy-transfer from tyrosine to tryptophan. The tryptophan fluorescence can be quenched not only by external agents (such as oxygen), but also by nearby residues. Furthermore, the environment of a tryptophan affects the position of its fluorescence spectrum maximum (which does not occur for tyrosine residues). Thus, variations in fluorescence intensity and/or wavelength of maximum intensity can be used to monitor stability, conformational rearrangements, dissociation, unfolding, and ligand binding and affinity of proteins and protein complexes, including viral particles.

Isolated sugars show no fluorescence, and the fluorescence work with lipids is rather limited due to the possibility of dispersion due to vesicle size. Puric and pyrimidinic bases in nucleic acids show a weak fluorescence (although they absorb strongly at 260 nm) at physiological pH, but they fluoresce strongly at acidic pHs and low temperature. Puric bases (especially guanine) show fluorescence at least three times more intense than pyrimidinic bases. In the native state, DNA or RNA show very weak fluorescence (except transfer RNAs which contains a highly fluorescent modified base); however, extrinsic fluorescent agents, that is binding to nucleic acids (such as ethidium bromide), can be used to detect nucleic acids. In this way, for example, the structural stability of some virions can be followed by quantifying the amount of nucleic acid released on capsid dissociation (and which is accessible to the external fluorophore).

Extrinsic fluorophores, such as ANS that binds to proteins, can change the protein fluorescence emission spectrum depending on the conformation of the macromolecule, and can be used to monitor unfolding, detect intermediate states or conformational rearrangements in proteins and protein complexes, including virus capsids. Furthermore, extrinsic fluorophores can be covalently attached to proteins, and used to map protein-ligand interactions within living systems [3, 4], after excitation at the appropriate wavelength.

6.2.2 Circular Dichroism

The Basis of CD

A property of many molecules is their chirality or molecular asymmetry. A description of the properties and characterization of organic chiral molecules can be found in organic chemistry texts [5]. The simplest cases result from the presence of asymmetric carbons; however, molecular chirality does not rely exclusively on the asymmetrically substituted carbon atoms, but rather in any molecular feature providing "handedness". For instance, proteins in the α-helical conformation wind in a right-handed manner; however, it is important to note that this helical "handedness" comes, at the last instance, from the stereochemistry of the amino acids. In nucleic acids, the helical polynucleotide winds in a left- or right-handed manner [6].

The chiral structures can be characterized by polarized light. In non-polarized light (Fig. 6.2a), the electric vector of the electromagnetic radiation can oscillate in any direction perpendicular to the way of propagation (for simplicity I shall reduce the following brief description to the electric field since, by using Maxwell's equations, we can characterize the electromagnetic radiation just by knowing the electrical component). Conversely, in a plane-polarized beam, the electric field oscillates only in a plane (Fig. 6.2b); if we moved with the field, we would see the vector representing the electric field oscillating back and forth along a line (Fig. 6.2b, right panel). In addition, in a circularly polarized beam, the single plane of oscillation rotates (Fig. 6.2c, left side): if we moved with the field, we would see the electric vector moving clockwise (right-handed circularly polarized light) or counter-clockwise (left-handed circularly polarized light). This single rotating plane can be decomposed in two components: the left and right circularly polarized vectors.

The two enantiomers of a chiral molecule have different refractive index, n; this phenomenon is called *birefringence*. That is, the variation of n is not the same through any direction in space in that material, and so, it is said that the variation of n is anisotropic. Biological cells or crystals have different n, depending of: (i) the direction of the beam, which illuminates them; and, (ii) the different planes of material they are made of. Chiral molecules show *circular* birefringence, and we say that the molecule shows circular dichroism. A good and rigorous description of the physical basis of CD can be found in several textbooks [7–9]. From a quantum perspective, CD is related to a transition between both electronic states in a molecule; this transition is active if, for an instant, the electric and magnetic components of both states are not perpendicular to each other (and this only occurs if the molecule is chiral). The CD spectrum results from the different absorption (or changes in the phase of the wave) of right-handed or left-handed circularly polarized light by a chiral molecule. In addition to de-phasing (relative to the incident beam), the amplitude of the emerging electric field is also modified differently for the left-handed and right-handed circularly polarized lights. Thus, the resulting electric field amplitude (the sum of the left-handed and right-handed circularly polarized electric field amplitudes) does not follow a circle, but it does an

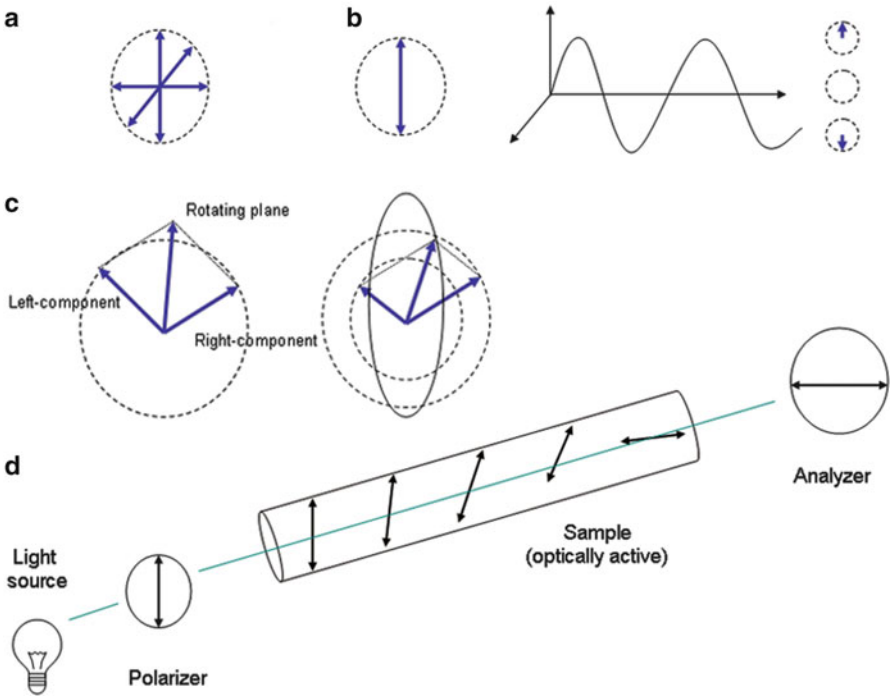


Fig. 6.2 Polarized light and CD. For simplicity, only the electric field (and not the magnetic field) of the electromagnetic radiation is shown (*double arrow head*): (**a**) Un-polarized light, as seen by a stationary observer. (**b**) Linearly (or plane) polarized light (*left panel*) as seen by a stationary observer; (*right panel*) the *small circles* show how the beam would be seen by an observer travelling with the beam in a linearly polarized light. (**c**) A circularly polarized light, as seen by a stationary observer, with the left and right circularly polarized components (*left panel*), which when added generate plane (linear) polarized radiation; (*right panel*) when the circularly polarized beam passes through a sample optically active, the left and right components have different amplitude and the resultant addition component is elliptically polarized (*continuous line*). (**d**) Set-up of a CD spectropolarimeter. The *green line* indicates the beam

ellipse (Fig. 6.2c, right side). It is said that the emerging beam is elliptically polarized, and then, the units of the CD are provided as a measured of the degree of the ellipticity of such ellipse. The changes in the ellipticity for a molecule depend on the wavelengths of the incident light and result in a CD spectrum (see examples in Fig. 6.3). There are specific rules able to predict which component of the polarized light is preferentially absorbed [9–11]. An experimental set-up for measuring CD in an optically active sample is shown in Fig. 6.2d.

CD of Biomolecules

Traditionally, the regions of the electromagnetic spectrum useful for CD analysis of biomolecules have been divided in: (i) the far-UV region (from 190 to 250 nm); and, (ii) the near-UV region (from 250 to 320 nm). In proteins, the peptide bonds

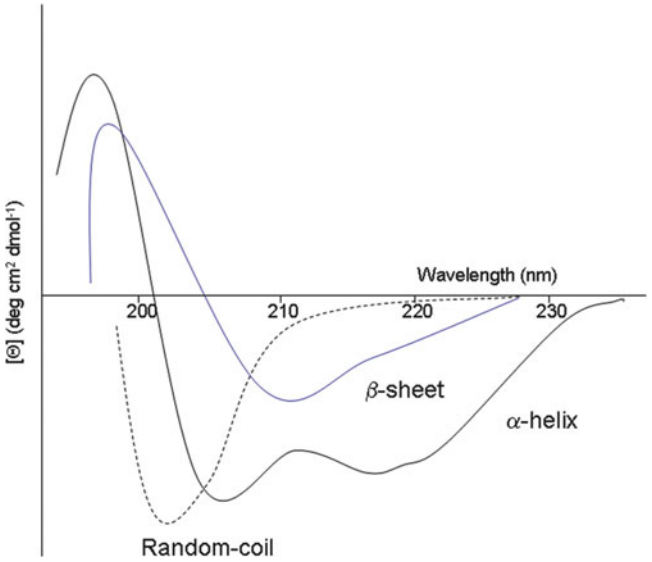


Fig. 6.3 Far-UV CD spectra of different model proteins: pure α-helix (*black continuous line*), pure β-sheet (*blue continuous line*), or random coil (*black dashed line*)

absorb UV light below 240 nm; the aromatic side chains absorb mainly in the 260–320 range (they can also absorb around 222 nm [11]); and the disulphide bridges show a broad absorption band at 260 nm. In addition, the spectra of the different types of secondary structure have characteristic shapes (see below). Some CD-based studies use wavelengths longer than those indicated above but, in most cases, they involve molecules with a strong coloured chromophore.

Secondary structure elements in proteins (α-helices, β-sheets or stretches with irregular structure such as loops or random-coil, denatured polypeptides) show a characteristic far-UV CD spectra (Fig. 6.3); perhaps the most characteristic is that of the α-helix, with two minima at 208 and 222 nm. Obviously, those proteins containing a mixture of the three types of structure will show a CD spectrum that is the resultant of the three spectra mentioned above, according to the percentages of the corresponding types of secondary structure elements. Deconvolution of the spectrum of a protein allows an estimation of the percentage of each type of secondary structure element, and there are programs implemented online to do so [12, 13]. Caution, of course, must be exercised (as with any mathematical procedure applied to complex experimental data), since the presence of different overlapping bands may lead to wrong conclusions on the secondary structure prediction of the protein.

The near-UV spectrum of proteins is usually rather weak, except for those with a large number of aromatic residues (thus, large protein concentrations in a wide path-length cell are generally required) [11]; contrary to what happens in the far-UV region, the near-UV spectrum of proteins is mainly associated with the

asymmetric environment of the aromatic residues. This region is only used in specific cases, since the information provided is similar to that obtained by fluorescence spectroscopy. On the other hand, the near-UV spectrum of double-stranded DNA is very intense due to the packing of the bases. There are also "model-spectra" for the different nucleic acid conformations [9, 11]. As it happens with proteins, the transitions among different conformers can be monitored by following changes in the spectrum. The CD analysis of lipids is rather limited, due to the possibility of dispersion because of the size of the vesicles formed.

6.2.3 Fluorescence and CD Spectroscopies to Study the Stability and Conformational Dynamics of Biomolecules

In addition to investigating some structural aspects of biomolecules and biomolecular complexes including viruses, fluorescence and CD spectroscopies can be used to detect conformational changes in these molecules and complexes by following changes in their fluorescence and/or CD spectra. One particular example is the determination of their conformational stability. As the far-UV CD spectrum of a random-coil is so different from those of the other types of secondary structure in proteins (Fig. 6.3), it is possible to follow the variations in the protein secondary structure by measuring the changes in the ellipticity at a selected wavelength (usually at 222 nm). Likewise, as the Trp or Tyr fluorescence in a protein is sensitive to changes in the exposure to solvent, it is possible to follow changes in the protein tertiary (and quaternary) structure by measuring the changes in fluorescence intensity at a selected wavelength (or, in the case of Trp, the changes in the wavelength of the maximum intensity). For the determination of protein stability by using these approaches, melting of the protein structure can be induced by a chemical denaturant (such as pH, urea or GdmCl) or temperature. In the simplest cases, if the melting is quite cooperative (an all-or-none transition) a sigmoidal curve will be obtained that can be fitted to a two-state model, and the thermodynamic parameters of interest can be obtained [14, 15].

6.3 Fluorescence and CD Spectroscopies to Study Virus Structure

CD and fluorescence spectroscopy have been used to study structural features, stability and conformational changes of isolated viral capsid proteins, the stable oligomers they form, and occasionally, even those of entire virus particles. In addition, both techniques have been used to study the interactions of viral proteins with other biomolecules. I shall divide the studies in the literature in each of those three different aspects.

6.3.1 Conformational Changes in Viral Proteins and Capsids

A general use of far-UV CD and fluorescence in structural virology is to investigate the overall secondary and tertiary structures of virus capsid proteins and their possible changes on capsid assembly. For example, the hepatitis B virus (HBV) capsid protein was found to be mainly helical, and showed little conformational restructuring on assembly [16]. There are many examples where mutations in the capsid proteins, which result in impaired self-association, do not lead to a significantly different fluorescence or far-UV spectra of the mutants, compared to those of the wild-type protein. In these cases, CD and/or fluorescence analyses showed that capsid assembly may proceed without large structural changes in the tertiary (or secondary) structures of the capsid building blocks; they also showed that subtle structural changes are enough to impair capsid assembly, which favours the possibility of developing antiviral strategies targeting this process (see Chap. 20). Conversely, dimerization of the capsid protein of human immunodeficiency virus type 1 (HIV-1) involves a transient monomeric intermediate and conformational rearrangements, although the overall fold of the monomer is maintained [17]. Changes in the structures of the capsid proteins occurred, for example, upon the pressure-induced dissociation of human rhinovirus (HRV14) [18], or the multimerization of the N-terminal region of Mason-Pfizer monkey virus p12 protein [19]. These and other examples using CD or fluorescence revealed that dissociation or association of some viral capsids may entail substantial conformational rearrangements, a fact that has been repeatedly confirmed and studied in structural detail using EM and X-ray crystallography (see Chaps. 3, 4 and 13).

Another general use of far-UV CD and fluorescence in structural virology is the study of conformational rearrangements of virus proteins in response to factors in the *in vivo* environment, by using agents which mimic *in vitro* the action of those factors within the host cell or virion. For instance, the HIV-1 polyprotein Gag contains a spacer polypeptide (SP1), whose first five residues are critical to allow the proper assembly of Gag into an immature capsid. The SP1 is unfolded in aqueous solution, but at high concentrations it shows a helix-like far-UV CD spectrum; thus, Gag aggregation possibly promotes the acquisition of secondary structure by SP1 which may mediate its function [20]. The effect of other intracellular or intravirion factors on virus protein conformation have been also mimicked *in vitro* by using salts (ions) at high concentrations, or organic solvents, changes in pH, or even by the addition of lipids; the conformational changes have been monitored by either CD and fluorescence. The results have contributed to reveal a substantial conformational dynamics of virus structural proteins, although no general trends are apparent, and the conclusions depend largely on the particular protein explored.

Likewise, far-UV CD and, more frequently, fluorescence spectroscopy have been used also to detect conformational rearrangements in assembled virus capsids, in response to factors such as heat or changes in pH that may mimic physiological conditions. As an example, fluorescence spectroscopy was used to detect a subtle conformational rearrangement in the capsid of the minute virus of mice (MVM) that was associated with a biologically relevant peptide translocation through capsid pores [21].

6.3.2 Stability and Association of Viral Capsid Proteins

Both CD and fluorescence spectroscopy have been used to determine the stability of viral proteins and particles in thermal or chemical denaturation assays (see Sect. 6.2.3). An excellent review on assembly as a binding process can be found in the work by Katen and Zlotnick [22] (see also Chap. 19; ref. [22] also contains an up-to-date review on the stability of several viruses followed by different spectroscopic techniques). For example, it was found that the wild-type HBV capsid protein dimer is highly stable, showing a T_m above 95 °C [16]. The ability of this dimer to self-assemble is improved by the naturally occurring mutation F97L, but this mutant has the same structure (as monitored by both techniques), and the same stability, as the wild-type [23]. The authors suggested that this mutation could affect the conformational dynamics of the dimer.

Among the comparatively few examples where the stability of whole virus particles was studied, urea-denaturation analyses of HRV14 followed by far-UV CD and fluorescence, showed a small stability [18]. However, the dissociated capsid proteins of HRV14, as analyzed by far-UV CD, had residual structure even at 8 M urea [18]. Interestingly, it was also possible to observe the signal in the near-UV of the viral RNA, and there was a decrease in the nucleic acid signal at 260 nm as the urea concentration was increased. This result suggests that the RNA becomes more ordered in the assembled capsid (see also Chap. 5).

Far-UV CD and fluorescence have been used also to identify the molecular determinants of folding, stability and oligomerization of capsid proteins by mutational analysis. For example, the thermodynamic roles of every residue at the dimerization interface in the HIV-1 capsid protein were probed by far-UV CD (secondary structure), fluorescence (tertiary structure) and gel filtration (quaternary structure) [24]. Mutational studies to analyze the folding, stability and/or oligomerization of viral capsid proteins have been reported also for bacteriophage P22 [25], *Physalis* mottle tymovirus [26] and MVM [27]. In all cases, thermal- and chemical-denaturations of the mutated residues, involved in inter-subunit interactions, disrupt both subunit protein folding and particle assembly. From these studies, it can be concluded that the distribution of critical residues and binding energy in intersubunit interfaces in virus capsids may be complex and quite different among viruses, and involve, in some cases, a few key residues, interactions and allosteric effects.

Studies on the folding, stability and association of viral proteins monitored by fluorescence and CD can be complemented by thermal denaturation analysis followed by differential scanning calorimetry (DSC); in this biophysical technique, instead of following a spectroscopic property, the heat exchanged in the dissociation and/or unfolding reaction of a viral capsid or protein is monitored. For instance, the thermal stability of the empty capsid of MVM have been studied by using fluorescence and DSC [21]. The DSC results indicated that the dissociation/denaturation transition involves a high enthalpy change and proceeds through one or more intermediates, which disappear in the presence of GdmCl. Furthermore,

thermal dissociation of the capsid involves a partially folded species of the capsid protein [27], with a high flexibility and practically devoid of tertiary structure. A similar partially folded conformation has been observed, by using fluorescence and CD, for the isolated capsid protein of herpes simplex virus (HSV) [28], and during maturation of the bacteriophage HK97 capsid [29]. These results, together with the existence of natively unfolded proteins (see Chap. 5), suggest that partially folded conformations (or completely disordered structures) of viral proteins may be very important in the virus life cycle [30]. Probably, flexibility may permit virus proteins to explore a wider conformational space, which facilitate the selection of the required conformation for self-association or interaction with other biomolecules (see Chap. 5).

The thermal stability of some viral capsid proteins have been analyzed at different pHs to elucidate the importance of several hydrogen-bonds, which disappear (or are formed) as the concentration of protons is modified. For instance, it has been shown that the thermal stability of the capsid protein of adenovirus type 2 is increased in mildly acidic conditions [31]. The effect of covalent modifications in viral capsid proteins on their thermal stability has also been analyzed. For example, PKC-mediated phosphorylation of the HBV core protein was shown to stabilize the protein, and also improved its self-association [32].

6.3.3 Binding of Ligands to Viral Proteins

If binding of a ligand induces a change in the secondary and/or the tertiary structure of the protein, the binding reaction can be monitored by far-UV CD and/or fluorescence, respectively. In a general sense, the addition of ions, or changes in pH can be considered particular examples of binding. Near-UV CD, together with fluorescence, can be used to detect the changes occurring in the protein (or the nucleic acid) upon binding of DNA or RNA. Alternatively, an increase in the T_m of the thermal denaturation of a protein monitored by far-UV CD, upon addition of a ligand [18] may indicate binding [33]: usually, upon binding, there is an increase in the T_m of the formed complex, when compared to that of the isolated protein. However, if the dissociation constant of the complex is large, or the heat released upon binding is very small, there might not be any changes in the measured T_m.

Both CD and fluorescence can be used to monitor the binding of ions to capsid proteins, which in some cases can trigger capsid assembly. A particularly well-studied example is that of HBV, where the affinity for Zn(II) has been determined by following changes in the fluorescence spectrum [22]. These findings suggested that: (i) the assembly reaction is kinetically trapped; and, (ii) the overall process of assembly triggered by the ion is allosterically regulated. Binding also can be also explored by using minimalist models of the capsid protein, as it has been described with fragments of human papilloma virus capsid [34]. Changes in fluorescence can be monitored not only on the capsid protein, but also on other molecules that can be used as reporters, as it has been shown,

for example, to demonstrate the inhibition of capsid assembly of *Salmonella typhimurium* phage P22 by 1,1-bis(4-anilinonaphtalene-5-sulfonic acid) [35]: the bisANS-bound coat or scaffolding proteins are incapable of assembly into procapsids.

Binding of antivirals or other molecules that affect virus capsid assembly (see Chap. 20) have been also monitored by fluorescence and/or CD [36]. In the cases reported, binding induces conformational changes in the protein, which can hamper proper assembly [37]; an excellent application of the design and use of these molecules can be found in the work by Zlotnick's group on HBV [38, 39].

6.3.4 Understanding Viruses: Some Major Contributions of Fluorescence and CD

The experiments with different viral systems, such as those described above, have proved the importance of spectroscopic analyses based on CD and fluorescence to analyze the folding, stability, conformational dynamics, and association of viral capsid proteins, and also to study the stability and conformational dynamics of viral particles, the effects of different factors, and the molecular determinants of these processes. The results using these and other biophysical techniques have revealed some rather general basic themes and a wide diversity of specific strategies followed by different viruses to respond to the selective pressures acting on the conformational stability and dynamics of capsids and virions.

6.4 Mass Spectrometry (MS) as an Analytical Tool

6.4.1 The Basis of MS

MS identifies molecules on the basis of their mass and charge and has become one of the most sensitive methods to study some aspects of virus structure. The first mass spectrometer was designed at the turn of the twentieth century, by Sir Joseph J. Thomson. Thomson observed the deflection of a beam of positively-charged ions, when subjected to a combination of magnetic and electrostatic fields. Ion deflection occurred through a series of parabolic curves, corresponding to particular mass/charge (m/z) ratios of the ions. The particular position of the curve was related to the velocity of the ion, and then, to its mass and acceleration. Although MS has evolved since those experiments, the basic set-up of the technique remains: all MS have three key elements: (i) an ion source; (ii) an analyser (that is, a sieve) of ion beams according to the m/z ratio; and, (iii) a detector monitoring the deflected beams (Fig. 6.4, top).

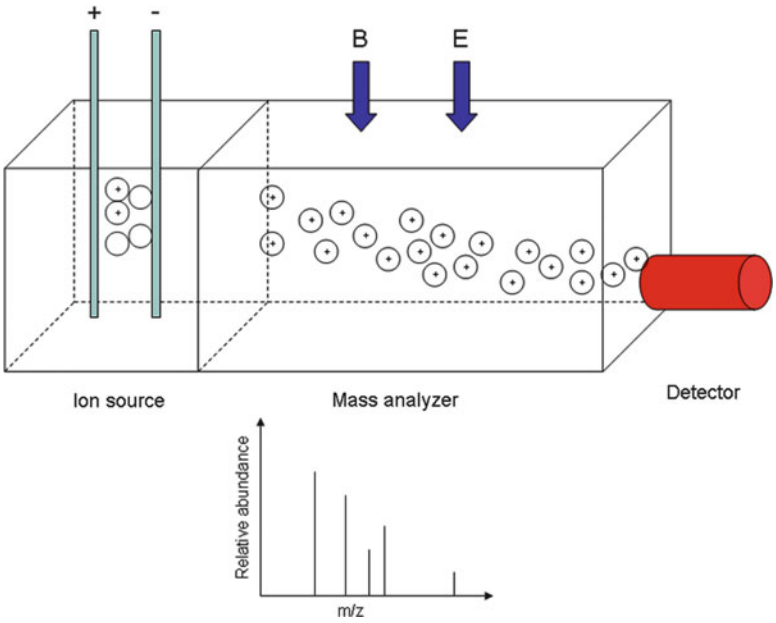


Fig. 6.4 Mass spectrometry. (*Top*) a mass spectrometer set-up, showing the ion source, the ion analyzer and the detector. In the ion analyzer, the deflection of the ion beams is shown. Initially, the sample is formed by neutral and charged molecules. (*Bottom*) example of a mass "spectrum" of a model protein

In the first stage in MS, molecules are vaporized and, in some experimental devices, also broken down (usually, but not exclusively, by the use of highly energetic particles); this molecular breaking produces ions. In the second stage, the ions are accelerated in a chamber where a high vacuum exists, under the influence of a combination of electric and magnetic fields; according to Newton's second law, lighter ions will deflect further than heavier ones, and they will follow different trajectories, depending on the ion masses and their charges. The third stage involves the monitoring of the ion deflections, which are used to calculate the m/z ratio, generating the mass finger-print, or the mass "spectrum". Graphically, a typical mass "spectrum" is a plot of intensity (signal) data *versus* the m/z ratio (Fig. 6.4, bottom) (and it is called "spectrum" in a wide sense, since no light is involved); thus, the nature of the mass "spectrum" will depend on how the spectrometer handles the sample: whereas some MS instruments break the molecules into several fragments, others record intact ones (that is, a single ion is produced).

6.4.2 Current MS Techniques

An excellent overview of the current methods in MS can be found in [40, 41], and its application to analyze protein complexes (focused in ESI-MS, see below) in

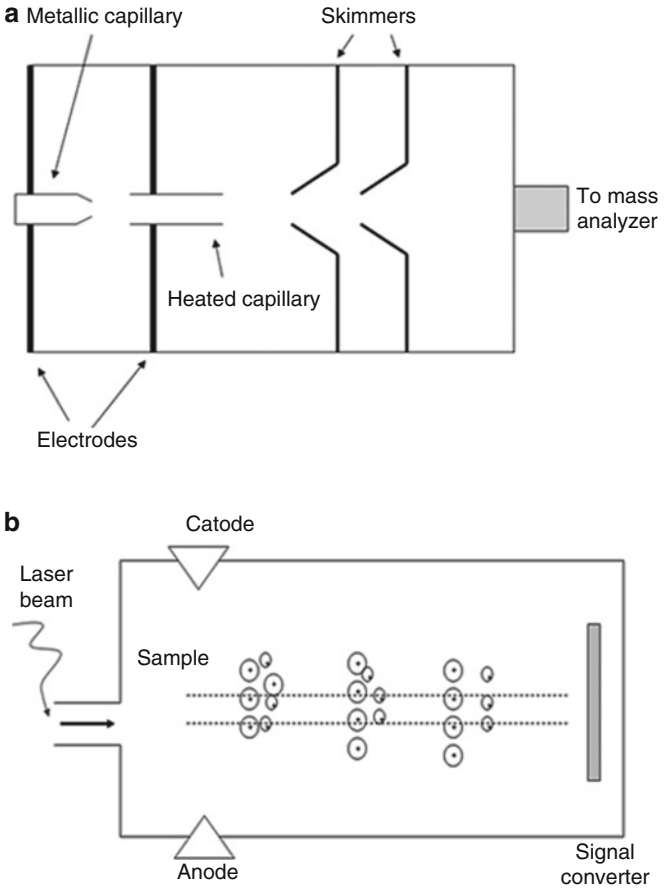


Fig. 6.5 Major types of MS in use. (**a**) ESI-MS schematic set-up representation: The solution is passed through a metallic capillary tube to produce a fine spray of droplets; the droplets cross through a large electric field which ionize them (*dark*, *thick lines* indicate the electrodes). The heated capillary dries the sample. The skimmers are cones with a central orifice designed to intercept the center of the droplet spray so as to sample the central portion of the expansion. (**b**) MALDI-TOF MS schematic set-up representation: The sample has been vaporized by the use of a laser beam, and it is ionized by the use of a electric field; as the ions pass through the analyzing chamber, the different ions are separated according to their sizes

[42]. There are currently two major MS techniques, which rely on different modes of sample ionization (in some examples, both techniques can be combined [42]):

(a) Electrospray ionization (*ESI*)-*MS*: An exhaustive description of the ESI technique can be found in [43, 44]. In this type of MS instrument, ESI is used to vaporize the sample. In ESI-MS, a volatile solvent with the molecule is pushed through a very small charged capillary, creating a spray of charged droplets, which during their evaporation leads to ionization (Fig. 6.5a), creating singly- or multiply-charged ions. The sample is directed towards the analyzer, which,

in most cases, is an ion trap. The ion trap uses a combination of magnetic and electric fields to trap ions in vacuum, by variation of the intensity and/or the direction of both fields (the so-called quadrupole trap).

(b) Matrix-assisted laser desorption ionization-time of flight (*MALDI-TOF*)-*MS*: This MS method is based on a soft ionization of the sample. This ionization uses a high-energy laser beam applied to a macromolecule, which is immersed in a protecting matrix (α-cyano-4-hydroxicynnaminic acid). The mixture is added to a plate and is allowed to dry (crystallize) before insertion into the MS. The laser causes the vaporization and ionization of the crystallized mixture (hence the name, matrix-assisted laser desorption ionization) (Fig. 6.5b). The ions produced are accelerated in the electric field in a vacuum long tube. We measure the time they take to reach the detector (their "time of flight"): larger ions reach the detector later than lighter (low m/z ratios) ones. As the acceleration occurs in vacuum, the velocity is wholly dependent on acceleration, which relies on mass, and ions reach the detector within 1 ms. MALDI-TOF-MS does not require sample fragmentation: it can be applied to biomolecules with a large mass, including viral particles; in contrast to ESI-MS, it relies on the production of singly-charged ions. Excellent reviews describing the advantages, applications and development of MALDI-TOF-MS can be found in [45, 46].

6.5 MS to Study Virus Structure

There are several advantages in the use of MS, when compared to other structural or analytical techniques described in this book (see Chaps. 3, 4 and 5, and Sects. 6.2 and 6.3). A first technical advantage is the very small amount of material required (usually two or three orders of magnitude lower than fluorescence or CD, and several orders of magnitude lower than for EM, X-ray crystallography and NMR). The second technical advantage relative to structural techniques is the speed of analysis, which can take only a few minutes. Regarding its uses in structural virology, MS is ideally suited to monitor changes and interactions of macromolecules in real-time: in particular, we can use MS to follow virus particle assembly pathways and conformational changes, obtaining information on the stability, dynamics and stoichiometry.

To study virus particles, MS is usually carried out at near-neutral pH, where ionization does not occur to a large extent, and relatively high ionic strength (close to physiological conditions). The low ionization has two major technical drawbacks: (i) weak signals in MS; and, (ii) high m/z ratios. Furthermore, the presence of salt and buffer ions (either specifically or non-specifically bound to the virus particles) leads to heterogeneous, broad signals, from which the mass cannot be easily and accurately obtained; for structural studies of virions containing nucleic acids the situation is worse, due to cation binding to the nucleic acid. Despite these drawbacks, MS is being increasingly used in varied and interesting applications in the field of virus assembly and structure. Several examples are provided in the three sub-sections below, which are interrelated.

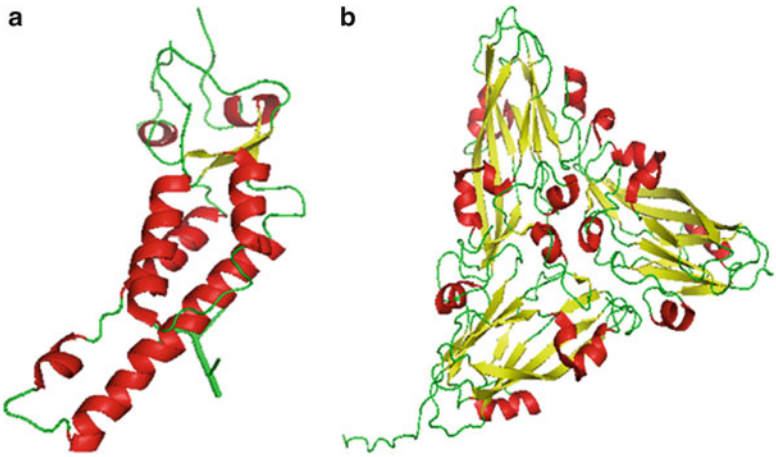


Fig. 6.6 Structures of two virus capsid proteins. (**a**) Structure of the TMV capsid protein (PDB number: 2OM3; [48]). The figure was obtained with PyMol [49]. A nucleotide-binding region is shown in *green* on the right side of the structure. (**b**) Structure of a trimer of the RYMV capsid protein (PDB number: 1F2N; [50]). Helices are shown in *red*, loop regions in *green* and β-strands in *yellow*

6.5.1 Virus Particle Composition and Structure

An early report of the study of an intact virus capsid by MS was described in the mid-nineties [47]. The author ionized rice yellow mottle virus (RYMV) and tobacco mosaic virus (TMV) (Fig. 6.6). Both are single-stranded RNA virus, but whereas RYMV has a 29-nm-diameter icosahedral capsid (with a total molar mass of 6.2 MDa), TMV forms a rod-shaped helical capsid with a diameter of 17.5 nm and a length of 300 nm (with a total mass of 40.5 MDa). The masses determined by MS were within ±15 % from the calculated ones [51]. Usually, in these experiments the samples originate high-charged states, with various m/z rates, which are difficult to assign. The authors showed that the samples used in the MS remained infectious and retained their native morphology; these experiments provided proof-of-concept for the non-invasiveness of MS, and the possibility of studying biological, weak, non-covalent complexes.

Since then, MS has been used to determine the masses of viral particles with different morphologies. For instance, the HBV capsids exhibit two different icosahedral geometries (T = 3 and T = 4, see Chap. 2), composed of 90 and 120 dimers of the capsid protein, and theoretical masses of ~3 and ~4 MDa, respectively. The experimentally determined masses were obtained by MS, and they were within 0.1 % of the calculated values [52]. These and other studies show that MS can be used to characterize virus particles by precise determination of their molecular weight.

6.5.2 *Virus Particle Stability, Conformational Dynamics and Assembly*

MS has been used also to explore the factors that determine the correct pathways of assembly, and stable enough assembly intermediates. For instance, the HBV capsid protein has two regions: a core (residues 1–140) and the protamine domain (residues 150–183), connected by a linker, whose exact length (and the conditions where assembly takes place) determines the proportion of T = 3 and T = 4 capsids that are formed during assembly (see above). MS has allowed to determine the rate of inter-conversion between both capsid forms, and that between those species and the soluble pool of capsid proteins (by using a mixture of ^{15}N-labelled with unlabelled proteins); therefore, the relative stability of both species can be also explored. Both capsids showed a similar temperature-dependent stability, although the T = 3 form maintained its integrity better during the MS experiments [52].

The stability and assembly of virus capsids may depend on capsid protein concentration, pH, redox conditions, the presence of disulphide-bridges, ionic-strength, metal ions, scaffolding proteins, the viral nucleic acid and other factors (see Chaps. 1, 10, 11, 12 and 19). A study on Norwalk virus provides an example on the use of MS to investigate the influence of pH on capsid assembly. The fully assembled capsid is formed by 180 copies of capsid protein VP1. MS analysis showed that the capsid is very stable at acidic pHs, but at alkaline pHs intermediate or aberrant structures formed by 2, 4, 6, 18, 40, 60 and 80 copies of VP1 were identified by MS (the latter two being the most abundant) [53]. These species could be reversibly converted into the native capsid (180 VP1 copies) at acidic pHs. The authors proposed a nucleation model of capsid assembly (see Chap. 19) based in the formation of smaller oligomers, mainly dimers, prior to their association into larger 60- and 80-copies intermediate species.

Another study on bacteriophage Qβ provides an example on the use of MS to analyze the influence of disulphide-bridges in capsid assembly [54]. MS analysis of capsid disassembly after disulfide bridge reduction yielded capsid protein dimers. However, disassembly of the intact capsids (without reduction of the disulphides) yielded mainly pentamers; reduction of the disulfide bridges in the pentamers led to monomers. Thus, the assembly pathway of the Qβ capsid may be directed by the intermolecular formation of covalent (disulphide) bonds around the capsid five-fold axes.

MS has been used also to study the influence of the viral nucleic acid on virus particle assembly (see also Chap. 12). Two examples are provided. In the first study, the *in vitro* reassembly of bacteriophage MS2 was triggered by a sequence-specific RNA-protein interaction between the capsid protein dimer and a RNA stem loop (SL) encompassing the Ψ-site (packaging signal) in the viral RNA [55–57]. MS analysis showed that the protein-RNA complexes had a 1:1 stoichiometry. In the second study, the concomitant dimerization and packaging of the HIV-1 genome (also analyzed by NMR, see Chap. 5) was investigated by using MS. The two copies of the genomic RNA are bound by an inter-molecular base-paired region: formed by the self-complementation of a sequence located in the SL of the Ψ-site. The

RNA dimer is stabilized through the interaction with the nucleocapsid protein, and the ESI-MS studies showed that binding-sites are present in two, and not all four, of the SLs [56]. Therefore, the comparison of the assembly pathways in the presence and in the absence of the SL for MS2 has allowed to discern the importance of the different intermediate self-associated capsid protein species to achieve an efficient capsid assembly [57]; in fact, intermediates in which each subunit protein capsid had only a handful of contacts with the nearby subunits would tend to be favoured during capsid assembly, and thus a nucleation mechanism was proposed.

6.5.3 Dissection of Conformational Stability and Dynamics in Virus Particles Using Limited Proteolysis-MS or Hydrogen-Exchange (HX)-MS

Major advances in the study of specific regions of a virus particle in its conformational stability and dynamics have been possible by using limited proteolysis or hydrogen exchange (HX) coupled to MS and a few other techniques. The interested reader can find this topic reviewed in [58, 59].

Limited Proteolysis-MS

Viral capsids are conformationally dynamic assemblies. Even at equilibrium, a capsid may sample different conformations and transiently expose some structural elements that would be otherwise buried (the so-called capsid breathing). Limited proteolysis-MS (and HX-MS, see below) has been used to identify dynamic regions in virus capsids. The limited proteolysis approach is based on the identification by MS of the peptides produced in a time-dependent, limited proteolytic attack on a virus particle. Particle regions that are conformationally more dynamic will, in general, expose cleavage sites more readily and will be cleaved earlier. The size of the fragments of the capsid proteins obtained at different times reveal the sites cleaved in the particle, and thus the dynamic regions [59].

By using this approach, the biologically relevant, transient externalization of (normally internal) terminal segments of the capsid proteins has been detected in picornaviruses (poliovirus, HRV) [60], nodaviruses (flock house virus, FHV) and HBV; in general, the tertiary and the quaternary structure of the virions can be mapped (see [61] and references therein). In addition, by using this approach, local and global breathing have been shown to differ between different states of a viral capsid, for example during maturation of tetraviruses (*Nudaurelia capensis* ω virus). The binding of the viral RNA genome to the capsid inner wall in FHV [62] or of hydrophobic compounds to pockets in the capsid of HRV [63] also impaired local or global breathing. Some of these effects on capsid dynamics have been shown to be important for virus biology (see Chap. 10) or the action of antiviral agents (see Chap. 20): whereas some anti-virals stabilized the capsids against proteolysis, others had the opposite effect (see [61] and references therein).

HX-MS

The basis of hydrogen exchange in biomolecules including proteins and protein complexes was explained in Chap. 5 as a part of the description of the HX-NMR technique. The pattern of exchanged protons with deuterium is dependent on the conformational dynamism of the region considered, and the number of incorporated deuterium nuclei in each region analyzed can be used to estimate its conformational dynamism. Since the mass of deuterium is larger (roughly twice) than that of a proton, the number of incorporated deuterium nuclei can be easily determined with precision by the current mass spectrometers [59, 64]. Several examples of the use of HX-MS to investigate the conformational stability or dynamics of virus particles, and specific regions in them are described next.

HX-MS studies on HRV indicated that intersubunit interfaces are conformationally rather static while regions around the capsid fivefold axes, involved in the translocation of polypeptide segments are more dynamic [65]. Some aspects of the conformational changes of the phage HK97 capsid during maturation (see Chap. 13) have been described by the combined use of protein engineering and HX-MS. During maturation, the HK97 capsid undergoes an expansion of roughly 20 % in diameter. The number of exchanged hydrogens suggested that a pre-expanded particle suffers significant refolding and twisting of the tertiary structure to accommodate the expansion [66, 67]. The differences in solvent-accessibility observed between capsid subunits in different quasi-equivalent conformations disappeared after the conformational transition that led to a first maturation intermediate. However, other quaternary interactions close to the icosahedral three-fold axes were preserved during maturation. These three-fold "staples" are formed by residues Arg, Glu and a metal binding site; mutations at those sites facilitated the conformational transition, as revealed by a larger proportion of protons being exchanged, and a decrease in the *in vitro* expansion rate. Taken together, these and other results suggested that the HK97 capsid, before the expansion, is in a high-energy metastable state.

The dynamics of maturation of the HIV-1 capsid has also been studied by HX-MS; both, the stability of the immature (either non-mutated or carrying a mutation that prevented the maturation cleavage) and the mature capsids have been described [68–70]. Among other results, it was found that the mature capsid is more stable (it exchanges more slowly) than the immature capsid. Changes in conformational dynamism during maturation and expansion of the bacteriophage P22 capsid have also been analyzed by HX-MS [71]. The P22 capsid protein consists of two domains connected by a stable linker. The HX pattern indicated that during maturation the C-terminal region maintained its conformational flexibility (dynamics), whereas the N-terminal region was made conformationally less dynamic. These experiments were consistent with the possibility that maturation increases the number of inter-subunit contacts.

HX-MS has been used also to investigate structural and dynamic aspects of nucleic acid encapsidation in double-stranded RNA or DNA bacteriophages. These phages rely on an active packaging mechanism, in which an ATP-dependent

molecular motor forces the nucleic acid into the preformed procapsid (see Chap. 12). These packaging motors consist of at least two regions: the ATPase complex and the connector portal complex; the portal complex replaces the capsid proteins at one of the five-fold vertices of the capsid. The use of HX-MS revealed, for example, intersubunit interactions between the portal and the capsid protein, and the number of subunits forming the portal (in P22, the portal when produced *in vitro* is a mixture of 11- and 12-mers) [70, 72, 73]. Conformational changes related with nucleic acid packaging have also been investigated by HX-MS. For instance, the structure of the packaging motor of the double-stranded RNA bacteriophage Φ8 was not structurally modified during ATP-binding and hydrolysis (as suggested by the absence of large changes in the HX pattern); conversely, RNA-binding caused changes in kinetics that allowed the identification of several nucleotide-binding sites [74]. Finally, the HX-MS results suggested than in the presence of the motor, the capsid is conformationally less dynamic [75].

6.5.4 Virus-Ligand Interactions

Certain MS conditions preserve non-covalently bound complexes including ligand-bound viral particles. Thus, the interactions between virus particles or their components and other biomolecules can be determined using MS. I shall describe several examples in the following paragraphs.

In several studies interactions between lipids or membrane proteins and viral capsids were detected by MS, and in some cases they were confirmed by other techniques such as EM (Chap. 3). For instance, the outer membrane proteins OmpA and OmpC protein from *Shigella*, the bacterial host of phage Sf6, have been detected in complex with the Sf6 capsid [76]. Also, MS in conjunction with enzyme digestion has shown that most of the proteins forming the capsid of iridovirus are able to bind lipids [77].

Some viral capsids recruit proteins from the host cells or other viral proteins required for infection. For instance, when HSV-1 nucleocapsids exit the nucleus, they become surrounded by a primary lipid envelope. The use of MS to detect mecromolecular interactions has allowed the identification of most of the proteins forming those primary enveloped virions: a glycoprotein, a cellular protein and a tegument protein [78, 79].

MS has also proved useful, together with electrophoresis, in analyzing the composition of viral particles and their covalent modifications. In many of these proteomic studies, cross-linking or proteolysis have been used to facilitate protein identification (see [61] for a review). These studies have helped in the determination of the stoichiometry of complexes involving viral proteins and/or nucleic acids. For instance, recently, a map of the interactions between HIV and human proteins complexes has been obtained; as many as 497 interactions, involving 435 human proteins, were identified [80].

6.5.5 *Understanding Viruses: Some Major Contributions of MS*

MS has become a very important technique for different analyses on the structure or properties of virus particles, as proven by the examples mentioned above and others. MS can be used to determine the size and composition of capsid building blocks, assembly intermediates, assembled capsids and virions, and aberrant viral particles. MS can be used also to follow structural changes in response to external factors, during assembly or disassembly of virus particles. MS has proven instrumental in the analysis of viral capsid breathing and in detecting conformationally dynamic regions in viral particles, and in the role of molecules or conditions facilitating or hampering those changes.

6.6 Perspectives

I foresee that CD, fluorescence spectroscopy and MS will hold a critical supporting role in structural virology. Fluorescence and CD are highly valuable for a first or rapid general structural characterization of viral proteins and viral particles; the information provided may prove particularly important in those cases where the protein or viral particle can not be crystallized, or can not be tackled by liquid NMR or solid-state NMR analyses. CD and, especially, fluorescence spectroscopy provide adequate probes for the rapid detection and following of some conformational rearrangements in virus particles. MS has become a critically important technique in several fields including structural virology. I anticipate that its use in this area will become widespread for the identification and study of the conformational dynamics of virus capsids, and for proteomic analyses in virology, including the detection of interactions between viral proteins and virus particles and cellular biomolecules which may participate in virus replication and assembly factories (Chap. 14). The power of the techniques described in this chapter is maximized in those cases where they can be combined (see Chap. 7) with structural techniques including EM (Chap. 3), X-ray crystallography (Chap. 4) and/or NMR (Chap. 5).

Acknowledgements I thank Luis Pérez and Francisco N. Barrera for careful reading of the manuscript, suggestions and new ideas. I apologize to those colleagues whose works were not cited due to lack of space or inadvertent omission. Work in the author's laboratory was supported partially by the Spanish Ministerio de Ciencia e Innovación (CTQ2011-24393 and CSD2008-00005), the FIPSE foundation (2006/0008), and intramural BIFI 2011 projects.

References and Further Reading

1. Lakowicz JR (1999) Principles of fluorescence spectroscopy, 2nd edn. Plenum, New York
2. Albani JR (2007) Principles and application of fluorescence spectroscopy. Blackwell, Oxford
3. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY (1995) Understanding, improving and using green fluorescent proteins. TIBS 20:448–455

4. Stepanenko OV, Stepanenko OV, Scherbakova DM, Kuznetsova IM, Turoverov KK (2011) Modern fluorescence proteins: from chromophore formation to novel intracellular applications. Biotechniques 51:313–327
5. Morrison RN, Boyd RN (1992) Organic chemistry, 6th edn. Prentice Hall, Englewood Cliffs
6. Wang AHJ, Quigley GJ, Koplpak FJ, Crawford JL, van Boom JH, van der Marel G, Rich A (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature 282:680–686
7. Schmidt W (2005) Optical spectroscopy in chemistry and life sciences. Wiley-VGH, Weinheim
8. Atkins P, de Paula J (2006) Physical chemistry for the life sciences. Oxford University Press, Oxford
9. Tinoco J, Sauer K, Wang Y (1996) Physical Chemistry: principles and applications for the life sciences, 3rd edn. Prentice Hall, Englewood Cliffs
10. Crabbe P (1967) Optical rotatory dispersion and optical circular dichroism in organic chemistry. Top Sterochem 1:93–98
11. Kelly SM, Price NC (2000) The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci 1:349–384
12. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 32: W668–W673
13. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89:392–400
14. Clarke J, Fersht AR (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. Biochemistry 32:4322–4329
15. Privalov PL (1992) Physical basis of the stability of the folded conformations of proteins. In: Creighton TE (ed) Protein folding. Freeman, New York
16. Wingfield PT, Stahl SJ, Williams RW, Steven AC (1995) Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. Biochemistry 34:4919–4932
17. Mateu MG (2002) Conformational stability of monomeric and dimeric forms of the C-terminal domain of human immunodeficiency virus-1 capsid protein. J Mol Biol 318:519–531
18. Gonçalves RB, Mendes YS, Soares MR, Katpally U, Smith TJ, Silva JL, Oliveira AC (2007) VP4 protein from human rinovirus 14 is elapsed by pressure and locked in the capsid by the antiviral compound WIN. J Mol Biol 366:295–306
19. Kneijzlik Z, Ulbrich P, Strohalm M, Lastůvková H, Kodícek M, Sakalian M, Ruml T (2009) Conformational changes of the N-terminal part of the mason-Pfizer monkey virus p12 protein during multimerization. Virology 393:168–176
20. Datta SA, Temeselew LG, Christ RM, Soheilian F, Kamata A, Mirro J, Harvin D, Nagashima K, Cachau RE, Rein A (2011) On the role of SP1 domain in HIV-1 particle assembly: a molecular switch? Virology 85:4111–4121
21. Carreira A, Menéndez M, Reguera J, Almendral JM, Mateu MG (2004) *In vitro* disassembly of a parvovirus capsid and effect on capsid stability of heterologous peptide insertions in surface loops. J Biol Chem 279:6517–6525
22. Katen S, Zlotnick A (2009) The thermodynamics of virus capsid assembly. Methods Enzymol 455:395–417
23. Ceres P, Stray SJ, Zlotnick A (2004) Hepatitis B virus capsid assembly is enhanced by naturally occurring mutation F97L. J Virol 78:9538–9543
24. del Alamo M, Neira JL, Mateu MG (2003) Thermodynamic dissection of a low affinity protein-protein interface involved in human immunodeficiency virus assembly. J Biol Chem 278:27923–27929
25. Green B, King J (1999) Folding and stability of mutant scaffolding proteins defective in P22 capsid assembly. J Biol Chem 274:16141–16146

26. Umahankar M, Murthy MRN, Savithiri HS (2003) Mutation of interfacial residues disrupts subunit folding and particle assembly of *Physalis* mottle tymovirus. J Biol Chem 278:6145–6152
27. Reguera J, Carreira A, Riolobos L, Almendral JM, Mateu MG (2004) Role of interfacial amino acid residues in assembly, stability and conformation of a spherical virus sample. Proc Natl Acad Sci U S A 101:2724–2729
28. Kirkitadze MD, Barlow PN, Price NC, Kelly SM, Boutell CJ, Rixon FJ, McClelland DA (1998) The herpes simplex virus protein, VP23, exists as a molten globule. J Virol 72:10066–10072
29. Lee KK, Gan L, Tsuruta H, Hendrix RW, Duda RL, Johnson JE (2004) Evidence that a local refolding event triggers maturation of HK97 bacteriophage capsid. J Mol Biol 340:419–433
30. Xue B, Willians RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN (2010) Viral disorder or disordered viruses: do viral proteins posses unique features? Protein Pept Lett 17:932–951
31. Rexroad J, Martin TT, McNeilly D, Godwin S, Middaugh CR (2006) Thermal stability of adenovirus type 2 as a function of pH. J Pharm Sci 95:1469–1479
32. Kang H, Yu J, Jung G (2008) Phosphorylation of hepatitis B virus core C-terminally truncated protein (Cp149) by PKC increases capsid assembly and stability. Biochem J 416:47–54
33. Waldron TT, Murphy KP (2003) Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. Biochemistry 42:5058–5064
34. Sun J, Yu JS, Jin S, Zha X, Wu Y, Yu Z (2010) Interactions of synthetic HPV-16 capsid peptides with heparin: thermodynamic parameters and binding mechanism. J Phys Chem B 114:9854–9861
35. Teschke CM, King J, Prevelige PE Jr (1993) Inhibition of viral capsid assembly by 1,1′-bis(4-anilinonaphthalene-5-sulfonic acid). Biochemistry 32:10658–10665
36. Zlotnich A, Mukhopadhyay S (2010) Virus assembly, allostery and antivirals. Trends Microbiol 19:14–23
37. Du S, Betts L, Yang R, Shi H, Concel J, Ahn J, Aiken C, Zhang P, Yeh JL (2011) Structure of the HIV-1 full-length capsid protein in a conformational trapped un assembled state induced by small-molecule binding. J Mol Biol 406:371–386
38. Zlotnick A, Lee A, Bourne CR, Johnson JM, Domanico PL, Stray SJ (2007) *In vitro* screening for molecules that affect virus capsid assembly (and other protein association reaction). Nat Protoc 2:490–498
39. Stray SJ, Johnson JM, Kopek BJ, Zlotnick A (2006) An *in vitro* fluorescence screen to identify antivirals that disrupt hepatitis B virus capsid proteins. Nat Biotechnol 24:358–362
40. Johnstone RAW, Rose ME (1996) Mass spectroscopy for chemist and biochemist. Cambridge University Press, Cambridge
41. Gault VA, McClenaghan NH (2009) Understanding bioanalytical chemistry: principles and applications. Wiley and Blackwell, Oxford
42. Ashcroft AE (2005) Recent developments in electrospray ionization mass spectrometry: non-covalently bound protein complexes. Nat Prod Rep 22:452–464
43. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246:64–71
44. Wilm MS, Mann M (1994) Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last? Int J Mass Spectrom Ion Proc 136:167–180
45. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal Chem 60:2299–2301
46. Tanaka K, Waki H, Ido Y, Akita S, Yoshida T (1988) Protein and polymer analysis up to m/z 100000 by laser ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2:151–153
47. Siuzdak G (1996) Mass spectrometry and viral analysis. Chem Biol 9:707–715
48. Sachse C, Chen JZ, Coureux PD, Stroupe ME, Fandrich M, Grigorieff N (2007) High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. J Mol Biol 371:812–835

49. DeLano WL (2002) The PyMOL molecular graphics system. De Lano Scientific, San Carlos
50. Qu C, Liljas L, Opalka N, Brugidou C, Yeager M, Beachy RN, Fauquet CM, Johnson JE, Lin T (2000) 3D domain swapping modulates the stability of members of an icosahedral virus group. Struct Fold Des 8:1095–1103
51. Fuesternau SD, Benner WH, Thomas JJ, Brugidou C, Bothner B, Siuzdak G (2001) Mass spectrometry of an intact virus. Angew Chem Int Ed 40:541–544
52. Uetrecht C, Versluis C, Watts NR, Roos WH, Wuite GJ, Wingfield PT, Steven AC, Heck AJ (2008) High resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. Proc Natl Acad Sci U S A 105:9216–9220
53. Shoemaker GK, van Dujin E, Crawford SE, Uetrecht C, Baclayon M, Roos WH, Wuite GJ, Estes MK, Prasad BV, Heck AJ (2010) Norwalk virus assembly and stability monitored by mass spectrometry. Mol Cell Proteomics 9:1742–1751
54. Ashcroft AE, Lago H, Macedo JM, Horn WT, Stonehouse NJ, Stockley PG (2005) Engineering thermal stability in RNA phage capsids *via* disulphide bonds. J Nanosci Nanotechnol 5:2034–2041
55. Stockley PG, Rolfsson O, Thompson GS, Basnak G, Francese S, Stonehouse NJ, Homans SW, Ashcroft AE (2007) A simple, RNA-mediated allosteric switch controls the pathway to formation of a T = 3 capsid. J Mol Biol 369:541–552
56. Fabris D, Chaudhari P, Hagan N, Turner K (2007) Functional investigation of retroviral protein ribonucleic acid complexes by nanospray Fourier transform ion cyclotron resonance mass spectrometry. Eur J Mass Spectrom 13:29–33
57. Knapman TW, Morton VL, Stonehouse NJ, Stockley PG, Ashcroft AE (2010) Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. Rapid Commun Mass Spectrom 24:3033–3042
58. Johnson JE (2003) Virus particle dynamics. Adv Protein Chem 64:197–218
59. Bothner B, Hilmer JK (2011) Probing viral capsids in solution. In: Agbandje-McKenna M, McKenna R (eds) Structural virology. RSC Publishing, London, pp 41–61
60. Bothner B, Dong XF, Bibbs L, Johnson JE, Siuzdak G (1998) Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. J Biol Chem 272:672–676
61. Morton VL, Stockley PG, Stonehouse NJ, Ashcroft AE (2008) Insights into virus capsid assembly from non-covalent mass spectrometry. Mass Spectrom Rev 27:575–595
62. Bothner B, Schneemann A, Marshall D, Reddy V, Johnson JE, Siuzdak G (1999) Crystallographically identical virus capsids display different properties in solution. Nat Struct Biol 6:114–116
63. Lewis JK, Bothner B, Smith TJ, Siuzdak G (1998) Antiviral agent blocks breathing of the common cold virus. Proc Natl Acad Sci U S A 95:6774–6778
64. Tuma R, Coward LU, Kirk MC, Barnes S, Prevelige PE Jr (2001) Hydrogen-deuterium exchange as a probe of folding and assembly in viral capsids. J Mol Biol 306:389–396
65. Wang L, Smith DL (2005) Capsid structure and dynamics of a human rhinovirus probed by hydrogen exchange mass spectrometry. Protein Sci 14:1661–1672
66. Gertsman I, Fu CY, Huang R, Kornives EA, Johnson JE (2010) Critical salt bridges guide capsid assembly, stability and maturation behavior in bacteriophage HK97. Mol Cell Proteomics 9:1752–1763
67. Gertsamn J, Komives EA, Johnson JE (2010) HK97 maturation studies by crystallography and H/^{2}H exchange reveals the structural basis for exothermic particle transitions. J Mol Biol 397:560–574
68. Monroe EB, Kang S, Kyere SK, Li R, Prevelige PE Jr (2010) Hydrogen/deuterium exchange analysis of HIV-1 capsid assembly and maturation. Structure 18:1483–1491
69. Cortines JR, Monroe EB, Kang S, Prevelige PE Jr (2011) A retroviral chimeric capsid protein reveals the role of the N-terminal β-hairpin in mature core assembly. J Mol Biol 410:641–652
70. Fu CY, Uetrecht C, Kang S, Morais MC, Heck AJ, Walter MR, Prevelige PE Jr (2010) A docking model based on mass spectrometric and biochemical data describes phage packaging motor incorporation. Mol Cell Proteomics 9:1764–1773

71. Kang S, Prevelige PE Jr (2005) Domain study of bacteriophage P22 coat protein and characterization of the capsid lattice transformation by hydrogen/deuterium exchange. J Mol Biol 347:935–948
72. Poliakov A, van Dujin E, Lander G, Fu CY, Johnson JE, Prevelige PE Jr, Heck AJ (2007) Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of the bacteriophage portal assemblies. J Struct Biol 157:371–383
73. Kang S, Poliakov A, Sexton R, Renfrow MB, Prevelige PE Jr (2008) Probing conserved helical modules of portal complexes by mass spectrometry-based hydrogen/deuterium exchange. J Mol Biol 381:772–784
74. Lisal J, Kainov DE, Lam TT, Emmett MR, Wei H, Gottlieb P, Marshall AG, Tuma R (2006) Interaction of packaging motor with the polymerase complex of dsRNA bacteriophage. Virology 351:73–79
75. Lisal J, Lam TT, Kainov DE, Emmett MR, Marshall AG, Tuma R (2005) Functional visualization of viral molecular motor by hydrogen deuterium exchange reveals transient states. Nat Struct Mol Biol 12:460–466
76. Zhao H, Sequeira RD, Galeva NA, Tang L (2011) The host outer membrane proteins OmpA and OmpC are associated with the *Shigella* phage Sf6 virion. Virology 409:319–327
77. Wu J, Chan R, Wenk MR, Hew C-L (2010) Lipidomic study of intracellular Singapore group iridovirus. Virology 399:248–256
78. Padula ME, Sydnor ML, Wilson DW (2009) Isolation and primary characterization of herpes simplex virus 1 primary enveloped virions from the perinuclear space. J Virol 83:4757–4765
79. Harper AL, Meckes DG Jr, Marsh JA, Ward MD, Yeh PC, Baird NL, Wilson CB, Semmes OJ, Wills JW (2010) Interaction domain of the UL16 and Ul21 tegument proteins of herpes simplex virus. J Virol 84:2963–2971
80. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K et al (2011) Global landscape of HIV-human protein complexes. Nature 481:365–370

Further Reading

Baker M (2010) Mass spectrometry for biologists. Nat Meth 7:157–161
Burlingame AL, Carr SA, Baldwin MA (eds) (2000) Mass spectrometry in biology. Humana Press, Totowa
Kelly SM, Ness TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta 1751:119–139

Also recommended for further reading are references [1, 2, 7–9, 40, 41, 43] *listed above*.

Chapter 7
Combined Approaches to Study Virus Structures

Daniel Badia-Martinez, Hanna M. Oksanen, David I. Stuart, and Nicola G.A. Abrescia

Abstract A virus particle must work as a safe box for protecting its genome, but at the same time it has to undergo dramatic conformational changes in order to preserve itself by propagating in a cell infection. Thus, viruses are miniaturized wonders whose structural complexity requires them to be investigated by a combination of different techniques that can tackle both static and dynamic processes. In this chapter we will illustrate how major structural techniques such as X-ray crystallography and electron microscopy have been and can be combined with other techniques to determine the structure of complex viruses. The power of these hybrid method approaches are revealed through the various examples provided.

Keywords Hybrid methods • X-ray crystallography • Electron microscopy • Cryo-electron microscopy • Electron tomography • Cryo-electron tomography • Small-angle X-ray scattering • Virus • Bacteriophage • Capsid • Mutagenesis • Dissociation • Crystal structure • Fitting

D. Badia-Martinez
Structural Biology Unit, CICbioGUNE, CIBERehd, Bizkaia Technology Park, 48160 Derio, Spain

H.M. Oksanen
Institute of Biotechnology and Department of Biosciences, Viikki Biocenter, University of Helsinki, P.O. Box 56, Viikinkaari 5, 00014 Helsinki, Finland

D.I. Stuart
Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK

Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Didcot, UK

N.G.A. Abrescia (✉)
Structural Biology Unit, CICbioGUNE, CIBERehd, Bizkaia Technology Park, 48160 Derio, Spain

Ikerbasque, Basque Foundation for Science, 48011 Bilbao, Spain
e-mail: nabrescia@cicbiogune.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_7,

Abbreviations

2D	Two dimensional
3D	Three dimensional
BMV	Bromegrass mosaic virus
EM	Electron microscopy
ET	Electron tomography
Fab	Antigen-binding antibody fragment
FMDV	Foot-and-mouth disease virus
HCV	Hepatitis C virus
HCV-LP	HCV-like particle
HIV	Human immunodeficiency virus
HRV-16	Human rhinovirus 16
HSV	Herpes simplex virus
MR	Molecular replacement
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
RVFV	Rift Valley fever virus
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
SBMV	Southern bean mosaic virus
SeMet	Seleno-methionine
SIV	Simian immunodeficiency virus
TBSV	Tomato bushy stunt virus
TEM	Transmission electron microscopy
TMV	Tobacco mosaic virus
TNV	Tobacco necrosis virus
VLP	Virus-like particle
WNV	West Nile virus

7.1 Introduction: The 'Multi-Disciplinary Approach' Concept in Structural Virology

7.1.1 Early Structural Studies on Viruses

In previous chapters several independent approaches to study virus structure including electron microscopy (EM, Chap. 3), X-ray crystallography (Chap. 4), nuclear magnetic resonance spectroscopy (NMR, Chap. 5), mass spectrometry (MS) and other spectroscopic methods (Chap. 6) have been described. In this chapter, we will illustrate how some of those methods can be combined to provide further insights into virus structures. To clarify how these initially independent methods have converged, complemented and combined with each

other in the study of viruses, we will provide a brief historical perspective and a recapitulation of some of these structural methods as applied to virology (Fig. 7.1).

Structural virology began when Ernst Ruska built the first electron microscope and together with his brother, medical doctor, Helmut Ruska initiated the imaging of viruses and other submicroscopic structures during the years 1938 and 1939. Alongside this, the principles for the small-angle X-ray scattering (diffraction) (SAXS) technique were proposed by André Guinier (see Sect. 7.2.2). In 1935, Wendell Stanley provided the conditions for crystallization of the first virus, the tobacco mosaic virus (TMV), a plant-infecting virus. For his studies in producing pure enzymes and viruses, in 1946 he was awarded with the Nobel Prize in Chemistry. Also Ruska received the Nobel Prize but in Physics for his fundamental contribution in electron optics, incredibly, 55 years after building of his first microscope! In 1937, Bawden and Pirie crystallized another plant virus, tomato bushy stunt virus (TBSV) but it wasn't until 1941 that crystals of TBSV and TMV produced clear X-ray diffraction patterns.

The 1950s saw the beginning of the modern biology with the determination of the three-dimensional (3D) structure of the DNA by James Watson and Francis Crick and the first high-resolution X-ray structures of proteins, myoglobin by John Kendrew and hemoglobin by Max Perutz. Interestingly, in 1956 Watson and Crick also proposed the principles of virus organization. According to their hypothesis based on electron micrographs of plant viruses and their propensity to form crystals and concepts of genetic economy, simple viruses were formed by capsid of multiple copies of relatively small proteins arranged in a symmetrical fashion, providing identical protein contacts among their neighbours. They suggested that a cubic symmetry or a cylindrical shell would satisfy the above requirements whilst allowing a larger volume of nucleic acid to be packaged. The case for icosahedral symmetry of viruses was made by John Finch and Aaron Klug in 1959 in unveiling the organisation of poliomyelitis virus by X-ray diffraction in 1959. Caspar and Klug continued on this idea and formulated the quasi-equivalence theory indicating that certain multiples of 60 subunits could also form viral capsids but the equivalent interactions among neighbour subunits could no longer exist. Klug's thinking on this was inspired by the work of Buckminster Fuller in particular by his book *The Dymaxion World of Buckminster Fuller*. Caspar and Klug based the quasi-equivalent theory on the fact that the capsomers are adaptable molecules that organize in different ways within a highly ordered structure thus introducing the concept of protein flexibility to explain these different arrangements (see Chap. 2). Methods development for 3D reconstruction from electron micrographs began in 1964 when Aaron Klug and Jacob Berger introduced the use of Fourier transform in micrograph analysis. This analysis was extended to object with rotational symmetry, helical symmetry (for rod-shaped viruses) and finally to icosahedral specimens (for an historical account see 'Further reading' at the end of this chapter). In the same period Walter Hoppe, who contributed to the development of methods and instrumentation in X-ray analysis moved to the electron microscopy field transferring the same momentum and know-how in the advances of electron microscopy methods (see the ptychography technique in Sect. 7.7.2).

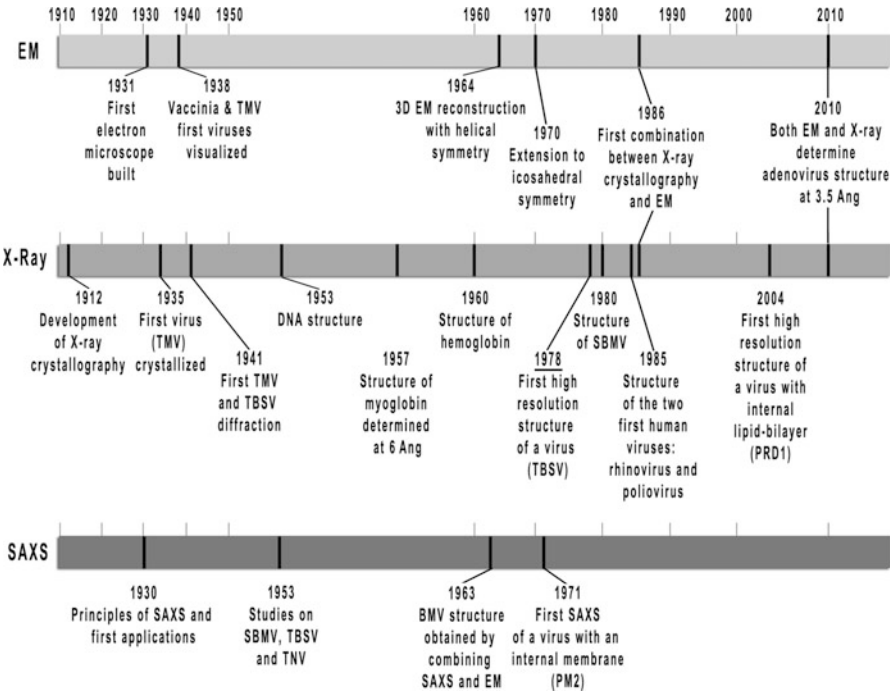


Fig. 7.1 Timeline of the major achievements in structural virology by EM, X-ray crystallography and SAXS techniques for the determination of virus structures. Due to their relevance the DNA and myo-hemoglobin structures have been included. Tobacco mosaic virus (*TMV*); Tobacco necrosis virus (*TNV*); tomato bushy stunt virus (*TBSV*); lipid-containing bacteriophage PRD1 (*PRD1*); southern bean mosaic virus (*SBMV*); bromegrass mosaic virus (*BMV*); lipid-containing bacteriophage PM2 (*PM2*)

Together with X-ray and EM, also small-angle scattering (using both X-rays [SAXS], and neutrons, [small-angle neutron scattering, or SANS]) was contributing to the analysis of virus structures. However in the years that followed, the impact of SAXS as a primary method became weaker leaving the leading role in structural virology to the other two structural techniques (Fig. 7.1). In parallel to the above developments, efforts were made to improve the diffraction of virus crystals and to devise ways to analyse symmetric structures that ultimately culminated into the structure of TBSV at 2.9-Å resolution in 1978 [1] and two years later into the structure of southern bean mosaic virus (SBMV) at 2.8-Å resolution [2]. Since then, both virus crystallography and electron microscopy have been fundamental techniques in providing structural insights in complex viruses. The techniques are complementary and well suited to hybridisation.

7.1.2 The Emerging Concept of Hybrid Methods

To grasp the concept of hybrid methods in structural virology one has to go back to the early culture of structural biology. The methods were developed to answer

fundamental biological questions but the X-ray diffractometers and microscopes were designed, built and maintained by engineers, physicists and biophysicists. The interdisciplinary expertises gathered in pioneering laboratories, in, among others, Birbeck and King's College in London (JD Bernal and Rosalind Franklin), Oxford University (Dorothy Hodgkin), the Royal Institution London (Lawrence Bragg and David Phillips) and the MRC Unit for the Study of the Molecular Structure of Biological Systems at Cambridge University (Max Perutz) around the middle of the twentieth century forged the spirit and the ambition to tackle the unexplored complexity of biology by complementary approaches (see Max Perutz's video interview and commentary on: http://www.youtube.com/watch?v=5I4eq63-3_I).

Thus, from one side, complex biological systems were reduced to smaller pieces (proteins or protein domains), in a reductionist (or "top-to-bottom") approach, studied separately and then, possibly, put back again in an overall framework (see also Chap. 5 regarding this approach in virus structure studies using NMR). On the other side, the complexity of larger macromolecular assemblies such as viruses could be looked at head-on thanks to the diverse expertise in those laboratories. So, hybrid methods and expertises were from the start supporting each other, validating the results obtained by the individual techniques.

Over the past three decades the multi-disciplinary approach has moved on from its original rationale towards the use of hybrid methods in predicting biological interactions, in uncovering dynamic processes and in bridging different resolution scales in molecular and cellular processes. The study of human adenovirus, a complex non-membrane icosahedral dsDNA virus (see below and Chap. 11) provides a good example of this. Indeed, this is the first case of combining X-ray crystallography and electron microscopy data, where the crystal structure of the viral capsid protein, assembled as stable capsomeric hexamers (hexons), was used to build an atomic model of the entire virion by exploiting the capsomer arrangement revealed by electron microscopy [3]. Since then the individual methods have become more powerful and the combination of (cryo-)EM and X-ray crystallography is a method of choice not just for many viruses but generally within structural biology.

In this chapter we will focus on describing how hybrid methods, with emphasis on the combination of X-ray crystallography and EM, have shaped our understanding of virus structures. We will provide examples of their contributions to the structure solution of very large and complex viruses and briefly discuss the challenges ahead, highlighting the power and vigour of methodological interactions in structural virology.

7.2 Some Classical Methods in Structural Virology: A Brief Overview

In this section we will recapitulate the three major structural techniques that have primarily contributed to the understanding of virus structures: EM, SAXS/SANS and X-ray crystallography. EM and X-ray crystallography as applied to viruses have been described extensively in Chaps. 3 and 4, respectively. Here we simply underline those concepts that link the methods.

7.2.1 Electron Microscopy

Transmission electron microscopy (TEM) relies on an electron beam passing through a virus sample either negatively stained to enhance the contrast or (in cryo-EM) vitrified at <100 K to ameliorate the radiation damage (vitrification is usually through simple flash freezing, but freezing rates may be enhanced by a simultaneous pressure wave; see Chap. 3). Electrons scattered elastically *via* the Coulomb potential of the atoms composing the virus (or any biological/material sample) are combined *via* magnetic lenses to generate a magnified image of the specimen. Depending on the TEM set-up data can be recorded in real space (images) or a reciprocal space (diffraction). In mathematical terms what allows the scattering process to be formalized in one or the other space is the Fourier transform (see also Chaps. 3 and 4). Since the energy of the electrons is very high their equivalent wavelength is much shorter than that for the X-rays used in crystallography. As a consequence in the case of image-mode, the result is a two-dimensional (2D) projection (real space) conceptually equivalent to the optical image obtained by cameras and visible light microscopes. In reciprocal space this corresponds to a plane of information which passes through the origin of reciprocal space (a central section) and information on both amplitude and phase of the transform of the elastically scattered electrons can be derived (Fig. 7.2). In the diffraction-mode only the amplitudes can be directly calculated from the measured intensity of diffraction spots (reciprocal space) but with higher precision and resolution than the corresponding values extracted from real-space images, which provide the less precise phase information (Fig. 7.2).

The dual-use of TEM in image- and diffraction- modes on the same sample constitutes the field of 2D electron crystallography. However, to collect electron diffraction images it is fundamental that the specimen forms a 2D ordered lattice, a requirement not needed for imaging mode. By combining these modes, 2D-electron crystallography circumvents the 'phase-problem' that X-ray crystallography must inevitably face (see Chap. 4). Each EM snapshot is equivalent to a 'still' X-ray image. Interestingly, it has proved much easier to generate 3D crystals than 2D (perhaps especially so for viruses which are often isometric), so the field of 2D crystallography remains much less developed.

Viruses have been investigated by EM since its early days due to their large and regular shape and relative ease of purification. Indeed, since the visualization of the first viruses in 1938–39 [4], a head-to-head race began in the technological advances of each of the major structural techniques at the time: EM, SAXS and X-ray crystallography (Figs. 7.1 and 7.2). Electron microscopy remains a workhorse method in several different aspects of virology from virus detection and identification to the 3D structure determination and analysis of virus-cell interactions. Continuing developments mean that nowadays electron microscopy can provide near-atomic resolution structures of viruses where atoms though not resolved, can be inferred [5] (see Chap. 3). The highest resolution analyses come from combining many projections of the virion collected from different particles thrown down at random;

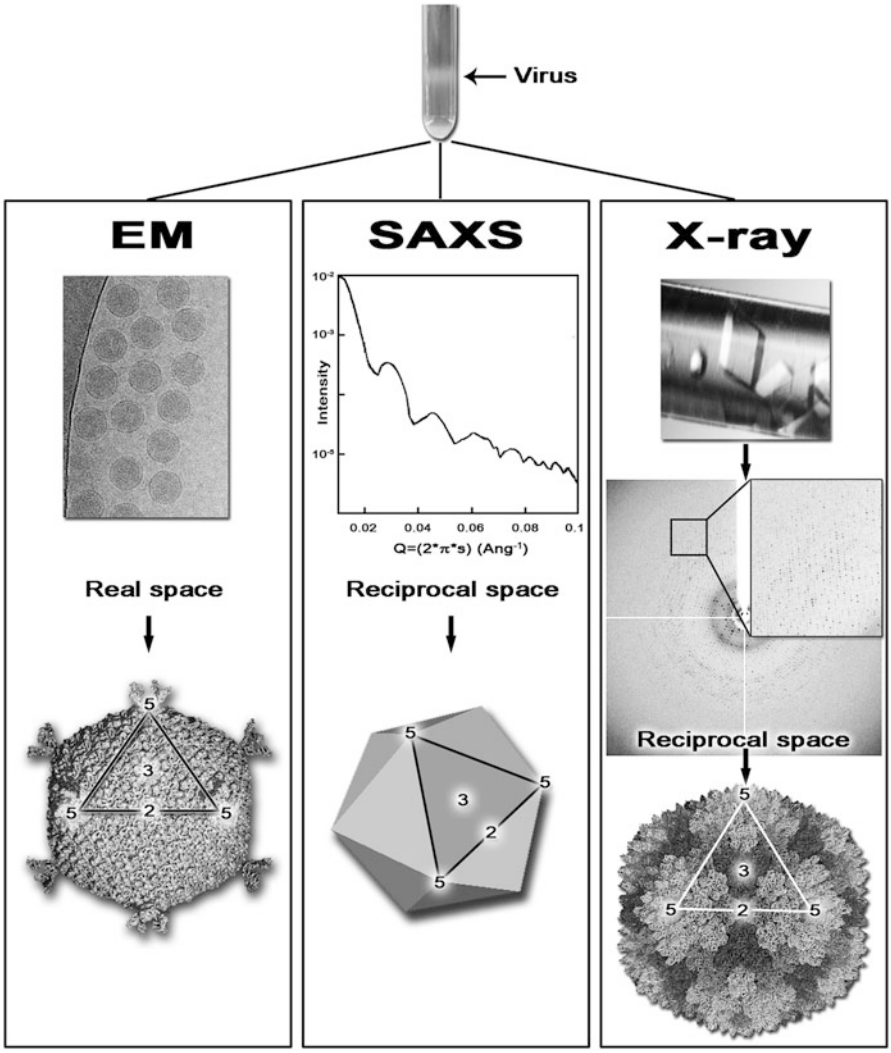


Fig. 7.2 Three major structural techniques at glance. Schematic representation of the workflow of virus structure determination in EM, SAXS and X-ray. All three techniques require purified virus; at the top, a *black arrow* indicates virus band resulting from ultracentrifugation. In EM the purified virus is placed on a microscopy grid, flash-frozen and then the 2D projections of the virus are visualized in an electron microscope producing real space images (*top*); post-processing of these images provides the virus structure (*bottom*). In SAXS the purified virus in solution is irradiated by an X-ray beam producing a low-angle scattering curve (*top*) containing raw data in reciprocal space; post-processing of the data produces the virus overall shape and molecular architecture. In X-ray crystallography the purified virus is used to obtain virus crystals (*top*) that are then diffracted producing diffraction images (*centre*) containing raw data in reciprocal space; once the data processing has been completed and the phase-problem solved, the virus structure is obtained at atomic resolution (*bottom*). The outlined region in the virus models in EM, SAXS and X-ray panels delineates delineates one of the 20 viral facets, and numbers indicate icosahedral symmetry axes (two, 2-fold; three, 3-fold; five, 5-fold). Each technique provides structural information in a different resolution range; in some cases these ranges overlap (see Fig. 7.4)

however in recent years it has become possible to get useful but lower resolution (perhaps 30-Å) data from an individual virus particle by collecting a full tilt series and doing a tomographic reconstruction (electron tomography, ET) (Chap. 3). Furthermore the resolution of both of these methods (EM and ET) can be enhanced for symmetrical virus particles by applying symmetry constraints (see Sect. 7.5.2). X-ray and electron methods begin to almost fully overlap when we consider X-ray microscopy, an emerging technique which is beyond the scope of this review to discuss, but it can simply be seen as a lower resolution, higher contrast version of ET capable of visualising specimens as thick as a eukaryotic cell [6]. Together this armoury of methods allows, for instance, the visualization of viruses and/or replication factories in-situ within the cell (see below and Chap. 14).

7.2.2 *Small-Angle X-ray Scattering*

SAXS is a technique that provides a structural characterization of a protein, protein complex or virus by using the elastic X-ray scattering from a solution containing the system of interest. The mathematical principles of small angle scattering were laid down in 1931 by André Guinier (Fig. 7.1) [for a review see [7]]. As used for the analysis of virus structure it is usual to work in solution phase and to use an X-ray (SAXS, wavelength usually ~1.5-Å) or neutron beam (SANS) onto a homogenous solution containing the biological specimen and to record the one-dimensional plot of scattered beam as a function of scattering angle (Q in the terminology of this method) close to the undeflected beam leaving the sample solution (hence "small-angle"). This one-dimensional scattering profile is the average of all scattering profiles deriving from the individual molecules (Fig. 7.2). Thus, in comparison to EM, where individual slices of the virus particle transform are assembled to produce as complete as possible version of the 3D transform, in small angle scattering the virus particle transform is spherically averaged to produce the linear scattering profile. Despite this massive loss of data, it is possible to extract information that relates to the volume, radius of gyration and the shape of the biological object. Modern synchrotron radiation sources are capable of producing highly acute curves which can support precise interpretation. Due to the development of sophisticated software which uses the massive computational power available today, it is, remarkably, possible to gain *ab-initio* low-resolution structural information by the use of the randomly distributed dummy atom modelling approach that has replaced the original trial-end-error shape assignment [8]. In brief, a large number of differently shaped molecules are generated and the scattering patterns simulated by analytical calculation are correlated and compared with the experimentally measured profile. Additional and/or external knowledge on the sample, for example disulfide-bonds presence, oligomerization state etc., will put constraints on the generation of the different models thereby increasing the reliability of the final shape assigned by the *ab-initio* procedure. Nonetheless, it is not possible to reconstruct a detailed 3D structure of a biological macromolecule from solely SAXS data due to the limited dimensionality of the raw data recorded (Fig. 7.2).

SAXS has been used in the characterization of the shape and composition of viruses since the early days of structural virology. Leonard and co-workers in 1953 carried out SAXS studies on SBMV, TBSV and TNV, in the same year the structure of the DNA was deduced from X-ray diffraction images (Fig. 7.1). At that point no virus structure was solved at high-resolution, but TBSV and TMV had already been crystallized and showed striking diffraction patterns. It is worth mentioning along this line that though not strictly SAXS, wide-angle diffraction and the interpretation of X-ray scattering from a partially ordered specimen, often referred to fibre diffraction have been pivotal in the elucidation not only of the DNA structure but also of virus structures such as TMV or phage Pf1 (for an exhaustive review on fiber diffraction studies of filamentous viruses see [9]). The use of such ordered fibres for structure determination has declined and nowadays is a technique almost unused.

In the early days turnip yellow mosaic virus, cucumber mosaic virus and bromegrass mosaic virus (BMV) were also investigated by SAXS. Interestingly, in the BMV study SAXS data were used alongside negative-stain images of the virus obtained by EM to reveal the virus dimensions and the presence of a hollow structure in the centre of the particles. The small set of plant viruses already mentioned were the initial common targets for the different structural techniques. A deviation from this theme was the structural study of a marine bacteriophage, PM2, the first phage to have been described possessing an internal lipid-bilayer [10]. This painstaking work combined both SAXS and negative-stain EM methods to describe in detail the architectural organization and assembly of proteins and lipids within the phage. Finally, fourty years later the *quasi*-atomic model of the entire PM2 virion was revealed by X-ray crystallography [11]. In retrospect the value of the method is evident – where specific questions can be asked, such as the radius or particle shell thickness and it is possible to construct a unique model, the method can give highly accurate answers and distinguish between competing models. This is exemplified by the results from neutron diffraction.

As with SAXS, SANS has provided insightful information of virus structures exploiting the differential contrast generated by the different interaction of neutrons with the nuclei of water (^1H_2O) and deuterated water (^2H_2O) to spatially map the genome and lipids from the protein mass [7]. Indeed, the advantages of this systematic contrast matching lies in the fact that since the intensity pattern measured (see Fig. 7.2) is the spherically averaged Patterson function defined as the Fourier transform of the structure factors describing the particle embedded in the solvent (for more details on Patterson function see general references on X-ray crystallography provided in Chap. 4), whole sets of Patterson vectors are eliminated by matching out certain components – disturbing markedly the diffraction.

In recent years SAXS has been re-emerging as an important structural tool to study biological macromolecules thanks to synchrotron radiation but also to automation in sample preparation and new high-capabilities of the DECTRIS-Pilatus detectors (see for example some of the European initiatives at http://www.saxier.org/ and at http://www.biostruct-x.eu/). However, from the virus point of view major insights using SAXS should be expected in the study of kinetic and time-resolved processes of virus assembly and disassembly which have started to appear [12].

7.2.3 X-ray Crystallography

In contrast to TEM and SAXS, virus X-ray crystallography (Chap. 4) relies on viruses ordering themselves in a 3D lattice, a crystal (Fig. 7.2). Illuminating the crystal by X-rays then allows one to record the Fourier transform of the unique portion of the lattice (unit cell) and hence the transform of the virus particle. The transform can only be sampled at those points specified by the Bragg condition, however this still allows a fully detailed reconstruction to be produced if all of the unique Bragg spots are measured and phases can be determined (see Chap. 4). Radiation damage is ameliorated since it is spread across all of the particles in the illuminated portion of the X-ray beam – usually in excess of a billion particles. All viruses crystallized so far possess icosahedral symmetry [icosahedron: a platonic solid with 20 identical equilateral triangular faces, 30 edges and 12 vertices; see Chap. 2 for details]. Due to this regular isometric shape, crystallization is often not the bottleneck in a virus crystallographic project and this is particularly true for relatively small and biochemically stable viruses. Indeed, the stability of virus crystals decreases with the increase of the virus size and it is approximately proportional to the inverse of the square of the virus radius [13]. Limitations in the amount or heterogeneity of the purified virus are the most restrictive factors in a successful project. However, recent advances in nano-crystallization techniques and efficient data collection strategies have allowed the determination of virus structures using only approximately 30 μg of virus as in the case of equine rhinitis A virus [14]. Once the crystals are available and show useful X-ray diffraction patterns, data recording and processing, phase determination and refinement with subsequent model building and refinement lead to the final atomic virus structure (for details of the entire protocol see Chap. 4 and [13]).

Thus, X-ray crystallography allows the determination of the virus structure at the atomic level and has contributed and still contributes to a very large extent to the understanding of virus biology. Milestones in structural virology (Fig. 7.1) have been the TBSV structure, the first virus solved at 2.9-Å resolution by Harrison's group in 1978 [1] and the structure of SMBV solved at 2.8-Å resolution by Rossmann's group in 1980 [2]. Five years later the structure of two viruses infecting humans, poliovirus and the common cold virus were solved by X-ray crystallography [15, 16] (see Chap. 4). Since then more virus structures have been elucidated by this technique as witnessed by the number of entries in the Protein Data Bank (PDB). The maximum number of virus structures was deposited in 2000. After this, the rate has decreased. In spite of this trend, in 2004, a significant contribution in the field came from the study of bacteriophage PRD1. PRD1 is the first and the only virus with an internal lipid-bilayer so far solved by X-ray crystallography in atomic detail, revealing a general principle of virus assembly scalable to very large viruses [17, 18].

Finally, in 2010, after several years of parallel advances, both X-ray crystallography and electron microscopy stunningly converged in terms of resolution achieved (~3.5-Å) with the elucidation of the structure of adenovirus (~900-Å diameter), which is so far the largest structure to be successfully tackled by X-ray

crystallography [19, 20]. Meanwhile the methods for the determination of protein structure by X-ray crystallography have become much more routine, in part thanks to efforts to systematise all aspects of the process by structural genomics/proteomics activities [21]. Methodological improvements have also impacted whole virus studies for instance data collection directly from crystallisation plates should make structure determination more routine and safer [22, 23] and continued developments at synchrotrons, including microbeams [24] and pixel array detectors indicate the viability of the method. An area currently unexplored is the use of nanocrystals at highly powerful pulsed X-ray sources (the so-called X-ray free electron lasers). Recent work suggests that structure analysis might be possible from much smaller arrays of viruses than has been possible with more conventional synchrotron sources (see Sect. 7.6.1 and Chap. 4). Different public databases [see *e.g.* PDB http://www.rcsb.org/pdb/home/home.do, http://www.ebi.ac.uk/pdbe/emdb/ and http://viperdb.scripps.edu/] bring together all structural information available on studied viruses.

7.3 Combining X-ray Crystallography and Electron Microscopy

7.3.1 General Approach: The Beginning of Quasi-Atomic Models

Since early times structural virologists have tried to merge information and techniques to provide reliable insights into virus structures. Prior to 1978 (when X-ray virus crystallography first revealed the atomic details of a complete virus [1]), SAXS, EM and X-ray diffraction were usually combined to cross-validate the low resolution information gathered by each individual technique. Increasingly powerful electron microscopes and the definitive establishment of protein and virus X-ray crystallography as a high-resolution technique have made it possible to combine structural information at different resolution levels yielding *quasi*-atomic virus models capable of predicting molecular interactions within the virus or between virus particles and ligands, including virus-host cell recognition processes and virus-antibody interactions.

EM and cryo-EM allow visualization and 3D reconstruction of whole viruses and virus-ligand complexes that cannot be studied by X-ray crystallography. At the same time X-ray crystallography can frequently provide atomic snapshots of the composite pieces of the virus and, as in a 3D jigsaw, structural virologists collate the different pieces to produce the most accurate and precise representation of the virus and its interaction with other (macro)molecules of the outside world. This process defines the procedure of fitting X-ray protein structures, either viral proteins or ligands of the virus into density maps of viruses and virus-ligand complexes obtained by EM and cryo-EM. Despite the fundamental differences in the way electrons and X-rays interact with matter, the end result of 3D analyses in each is a

'density' map which looks surprisingly similar between the two methods and can therefore be interpreted and analysed in a similar way, using similar software. It is important to mention that NMR spectroscopy, although it cannot tackle a whole virus structure, has provided structures of viral proteins that have been subsequently used for fitting into EM maps (see Chap. 5). NMR is also increasing its role as hybrid method providing constraints on possible dynamic events or conformational changes of virus structural components (see Chap. 5).

After an initial manual docking of the entire viral and/or ligand protein structure into the corresponding volume of the EM reconstruction aided by graphic computer programs, improvement and refinement of the fitting carried out in either real space or reciprocal space, can be performed using a correlation coefficient-criterion and/or a least-square minimization [25]. As a result, this rigid-body refinement protocol aims to determine a total of six parameters, a rotation matrix and a translation vector to optimize the fitting to the virus map (Fig. 7.3a). It goes without saying that for a successful fitting the scale of the EM derived density map has to match that of the atomic model.

It has been estimated that this procedure can reach an accuracy of 2.2-Å in a virus map of 22-Å resolution and of 4-Å in a 24-Å map [27], the higher the resolution of the map the better the docking. Programs available for this task include the General Averaging Program (GAP; D.I. Stuart and J.M. Grimes, unpublished software available upon request from authors), EMfit (http://bilbo.bio.purdue.edu/~viruswww/Rossmann_home/softwares/emfit.php), Situs (http://situs.biomachina.org/index.html) and UROX (http://sites.google.com/site/xsiebert/urox). Recently, the concept of flexible fitting has also been introduced that allows distinct domains of the fitted structure to undergo conformational changes to match the volume restrictions imposed by the low-resolution EM map, cf. Norma (http://www.igs.cnrs-mrs.fr/elnemo/NORMA/) or Flex-EM (http://salilab.org/Flex-EM/). One of the critical points in flexible fitting is how to define what is fixed and what is flexible. This can take advantage either of prior knowledge or of normal mode analysis (*e.g.* Norma) that mathematically assigns directions of motion to structural elements of the structure to be fitted. Either approach to the definition of the rigid bodies has its own strengths and weaknesses, for example a user might tend to be over conservative and not reflect the dynamic properties of the structure shown by the EM reconstruction whilst with normal modes the assignment of too many motions can lead to over interpretation.

However, if the resolution of the virus map reaches 7-Å then it becomes possible to identify the fold of the most ordered part of the virus, cf. of major capsid proteins, and to model it using templates for β-barrel or α-helical motifs [28]. It is also possible to adopt template substitution methods, such as those provided by the Rosetta software [29] for modelling into low resolution X-ray and EM derived density maps (4–7 Å).

Nevertheless a major problem when combining different levels of information and specifically in generating (virus) quasi-atomic model is how to finely balance prior knowledge relative to the resolution achieved. In Fig. 7.4 we show the number of effective data points *vs* parameters for the different methods allowing the reader

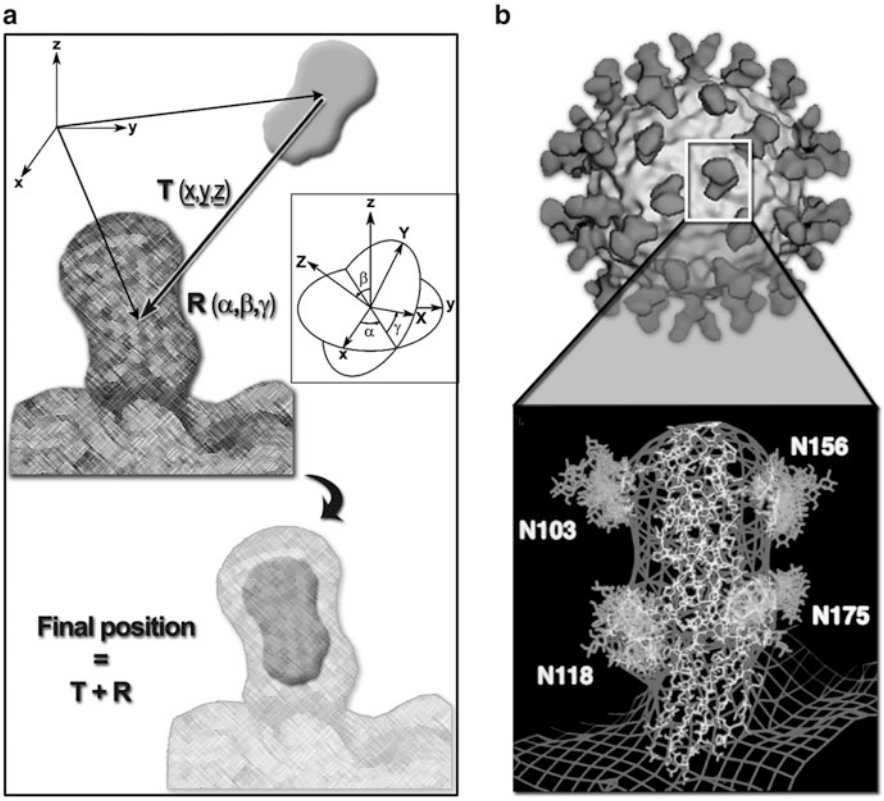


Fig. 7.3 Fitting X-ray data into EM maps. (**a**) Schematic of the fitting procedure; the crystal structure of a virus-binding molecule (ligand, represented here as a *grey* two-lobe object) is moved coarsely from its initial position into the corresponding electron density map of the virus-ligand complex obtained by EM (*dark-grey mesh*) by a translation (*T*) and a rotation (*R*); the final position of the object within the density is achieved by the refinement of six parameters [x, y, z and α, β, γ (Euler angles; inset)] using either real- or reciprocal-space refinement procedures (see [25]). (**b**) *Top*, EM reconstruction of rhinovirus 16 (*light-grey*) with CD155 cellular receptor molecules bound (*dark-grey*); *bottom*, the CD155 X-ray structure (*white* stick model) fitted into the virus-receptor EM density (*dark-grey* mesh); the N-glycosidic groups in the CD155 structure are labelled according to the asparagine residue to which each is bound (Adapted from [26]. With permission)

to grasp what is achievable by each technique and by their combination. In principle we need an objective way of assessing the information content of each of the sources of data and producing a model that satisfies the data to the level of accuracy of the data, but otherwise assumes nothing. We return to discuss this challenge below.

As mentioned earlier, one of the first examples of structural methods combination can be found in the study of adenovirus by Roger Burnett and his laboratory in 1986 [3]. Adenovirus is a large dsDNA virus that infects vertebrates, including

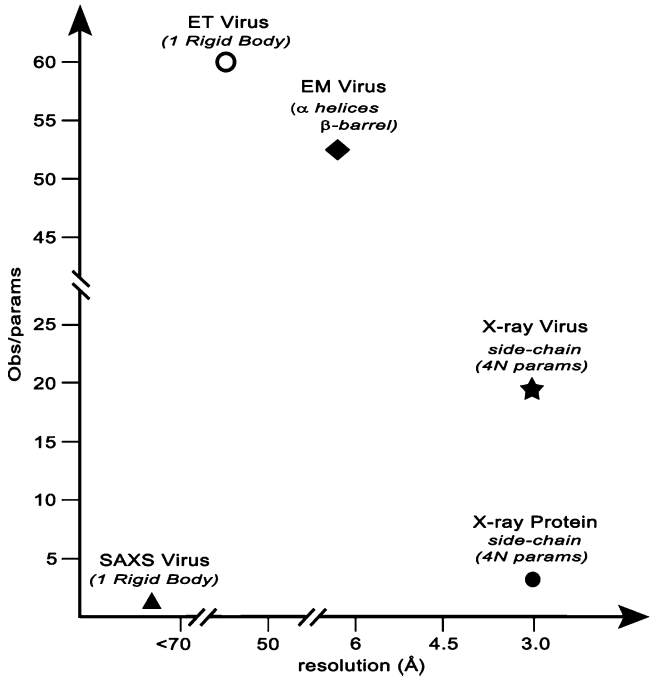


Fig. 7.4 Simplified ratio *per* technique. The data points were obtained by estimating the number of observations for a virus of a determined size and achievable resolution divided the number of structural elements (parameters) to be fitted. For the determination of the virus structure by X-ray crystallography virus we considered the lowest non-crystallographic symmetry scenario in a virus crystal

humans. As passionately stated by Roger Burnett (http://www.cicnetwork.es/secciones/interviews/11_burnett.html) his interest in adenovirus stemmed from beautiful electron microscopy images collected in the '70s by Nick Wrigley of hexamers of the major capsid protein, named hexons that recapitulated architectural elements of the entire virion. However, X-ray structural studies on the hexon structure had already started [30] prefiguring the approach that finally led to the interpretation of protein-protein interactions in the assembling of the large icosahedral adenovirus particle. By solving the X-ray structure of the hexon at 2.9-Å resolution and by placing it into the adenovirus EM reconstruction, Roberts and co-workers [3] opened the way to the hybrid methods approach in structural virology. This has been followed by many other studies for example of the binding of the human cellular receptor ICAM-1 to human rhinovirus 16 (HRV-16) [26], or neutralizing antibody bound to foot-mouth-disease virus (FMDV) [31].

More recently *quasi*-atomic models have been generated for enveloped and membrane-containing viruses. In the former case, a lipid-bilayer decorated by

glycoproteins surrounds a nucleocapsid whilst in the latter case the vesicle is engulfed by the proteinaceous capsid shell. One of the best structurally described enveloped virus is human pathogenic dengue virus (*Flaviviridae* family). Years of investigation on the entire virion by EM and of its glycoproteins by X-ray crystallography have finally produced a clear picture of the dengue virus maturation, assembly pathway and cell-entry mechanism [32, 33]. Also, by combining the crystal structure of an antigen-binding antibody fragment (Fab) in complex with West Nile virus (WNV) E glycoprotein and the dengue and WNV structures in complex with the Fab by EM, Cherrier and co-workers [34] have been able to establish the structural basis for antibody neutralization of infectious immature particles. Likewise *quasi*-atomic models for other biomedical relevant enveloped viruses have been recently generated with a similar 'fitting X-ray structures into cryo-EM maps' approach, for instance chikungunya virus (*Alphavirus* genus, *Togaviridae* family) and vesicular stomatitis virus (*Rhabdoviridae* family).

In the case of lipid-containing viruses, the original example in hybrid methods is PRD1, a bacteriophage that infects gram-negative bacteria [35] and whose *quasi*-atomic model was proposed in 2001 [36]. PRD1 is indeed the model system for this type of viruses having been intensively studied biochemically, genetically and structurally. Its architecture, major capsid protein fold and principle of assembly strikingly recapitulate those elucidated years earlier for adenovirus [3, 17, 18, 37] leading to the concept of viral lineages and virus common ancestry [38] (see below, Sect. 7.8). Another illustrative example, although not strictly a case of fitting X-ray structures into cryo-EM map, is presented by the process that led to the X-ray structure solution of the marine lipid-containing bacteriophage PM2. The moderate 7-Å resolution achieved with the PM2 virus crystals (in comparison to the ~4-Å resolution obtained for PRD1 virus) prompted the determination of the structures of the individual components at higher resolution [11]. This "divide et impera" approach led to high resolution structures of the two major viral capsid components that were then docked into the low-resolution X-ray map. As a result, key aspects of viral evolution, capsid assembly and membrane morphogenesis were elucidated [11].

Furthermore, it is worth underlining that the protocol of fitting of atomic structures into electron density maps can be also extended to volumes obtained by sub-tomogram averaging (see below and Chap. 3; sub-tomogram averaging algorithms [39–41] allow averaging equivalent structures thus improving the signal-to-noise ratio). This can lead to plausible fitting of X-ray crystal structures as in the case of the capsid protein domain of the Gag polyprotein in HIV [42].

Finally, the combination of methods in structural virology has pioneered an equivalent approach nowadays emerging in biology and defined as 'integrative modelling'. Indeed, this approach aims to combine different available sources of information – structural, biochemical, biophysical and proteomics – to obtain a molecular picture or a unified structural view of the system in study. Nice examples of this are embodied in the 'Integrative Modelling Platform' (IMP) (http://salilab.org/imp/) and Haddock (http://www.nmr.chem.uu.nl/haddock2.1/), with a recent example being the analysis of the nuclear pore [43]. How and with what weights the different sources of information are combined remain to be fully and routinely

implemented in a suitable statistical framework (Bayesian, Maximum-Likelihood, etc.) that would provide a validation and statistical significance assessment of the resultant model. Although this work is ongoing and is still very crude in many cases, the vision is already set, thanks to the combination of methods used to study virus structures.

7.3.2 *Validating the Quasi-Atomic Models*

Validation of the proposed *quasi*-atomic model is fundamental to the reliable interpretation of molecular interactions and their usefulness. Although several programs (see Sect. 7.3.1) that carry out positional refinement of the initial docked structure into low resolution maps provide correlation coefficients (for real space analysis), or metrics equivalent to the crystallographic R-factor [44] (for reciprocal space analysis), the inclusion of such values in the final publications is still left to the discretion of the authors. It must be said however that there is no unique criteria that describes the accuracy of the model and a multi-parameter assessment (atomic clashes, stereochemistry, biological meaningfulness) together with the statistical measures is the best solution at present. Furthermore, if additional biochemical or biophysical information is available, for instance a proved interaction between specific residues of the ligand and the virus, then it should be taken into account for example by imposing restraints on the movements of the fitted X-ray structure during refinement.

In 2010, the EM community attempted to standardize practices during the inaugural meeting of the Electron Microscopy Validation Task Force (http://vtf.emdatabank.org/). A series of guidelines and recommendations were suggested with the ultimate goal of increasing the impact and the significance of EM in biology and biomedicine. This reflected the earlier successful efforts of the X-ray community through the X-ray Validation Task Force (http://www.wwpdb.org/workshop/2011/index.html), where the broad principles are now accepted and required for publication. Given the limitations on the validation criteria it is advisable when combining X-ray with EM, to keep a conservative attitude in the biological interpretation of the resulting *quasi*-atomic model. Also, remember that the data-to-parameter ratio generally available during this hybridization process is unfavourable (see Fig. 7.4) – better safe than sorry!

7.3.3 *Phase Information "Interplay"*

A major problem in crystallography, as previously stated, is the so-called phase problem, a methodological 'inconvenience' that prevents the retrieval of atomic

positions directly from the raw data (see Chap. 4). EM images embody the phases of the object, by visualizing it in real space and thus are a potential source of phase information for diffraction methods. When highly symmetrical particles such as icosahedral viruses are investigated structurally, it is possible to exploit the redundancy of information for phase refinement (see Chaps. 3 and 4). In crystallography, in the best-case scenario when none of the symmetry axes of the virus align with the crystallographic symmetry axes of the crystal, the full 60-fold non-crystallographic symmetry (NCS) can be iteratively used to improve the starting phase assignment [in the worst-case scenario only 5-fold NCS]. It has been shown that the availability of 60-fold NCS can allow convergence of phase refinement starting from phases derived from a spherical object [45, 46] (for an overview on *ab-initio* phasing see [47]). However, for this protocol to succeed it is paramount to know (*i*) the exact location of the icosahedral symmetry axis relative to the asymmetric unit (AU) of the unit cell, (*ii*) the radius of the starting model [45, 46] and (*iii*) the protein region for the generation of appropriate envelopes used in the averaging and solvent flattening procedures.

In virus crystallography, apart from the first virus structures that were solved using heavy atom methods [1, 2], it has been and still is a common practice to use as a starting phasing model a virus previously solved [unrelated or even a sphere (see above)] or, if available, a low-resolution derived EM map located within the asymmetric unit of the unit cell of the crystal through the molecular replacement (MR) technique [13, 48] (see Chap. 4). It is worth noting that for the phasing of FMDV [49], a hybrid virus model was built from existing viruses weighting-up the conserved structural pieces and used as a searching model in MR, an approach that is now implemented in some of the standard protein MR protocols.

Advances in EM (see Chap. 3) have propelled the application of this phasing strategy to X-ray data obtained from crystals of biological macromolecules smaller than viruses and often less symmetrical [50–53]. Although there are exciting perspectives of this alternative approach for circumventing the crystallographic phase problem, further requirements need to be fulfilled for its successful application: resolution range overlap between the 3D electron microscopy data and those obtained by X-ray crystallography, *i.e.* it is important to collect the low-resolution terms, and the presence of non-crystallographic symmetry. The former condition allows the object to be located within the asymmetric unit of the unit cell (by MR), while the latter boosts the phase refinement and phase extension procedures. To be more precise the power of phase extension depends on the degree of oversampling of the transform of the unique portion of the object and non-crystallographic symmetry delivers this directly, however a very high solvent content does so as well (for instance a thin virus shell with disordered interior).

One of the first examples in which the use of low-resolution models was successful in phasing X-ray data of a small protein (38 kDa) was inspired by studies on Mengo and FMDV viruses [50]. Villeret and co-workers were not only able to build the correct geometry of the oligomeric state (12-fold) using a monomeric X-ray structure of a related protein and generate low-resolution phases at 8-Å but they were also able to phase-extend those initial phases to the full 3-Å resolution of the X-ray data (see also ‘Further reading’).

In principle also SAXS could provide initial low-resolution phasing as it has been described in the case study of nitrite reductase, a 34.4 kDa protein that forms homotrimers. However, in this case once the SAXS envelope was correctly located in the asymmetric unit of the unit cell, only phases to 20-Å were reliable and no further phase extension was attempted [54].

A recent enlightening case of phase interplay between EM and X-ray crystallography comes from the structure of the marine bacteriophage PM2 at 7-Å resolution [11] (Fig. 7.5). The initial phases for solving the entire virus X-ray crystal were derived from the cryo-EM based PM2 structure at 8.4-Å resolution [55]. In spite of the apparently minimal ~1.4-Å improvement in resolution of the X-ray over the EM data, the quality of the resulting X-ray map thanks to the accuracy of the crystallographically measured structure factor amplitudes, allowed a clearer interpretation of the different, 60-fold symmetrical arranged structural components, in particular those pivotal elements located between the capsid and the lipid vesicle [11]. However, the 7-Å X-ray map could not provide the unequivocal determination of the fold of the major capsid protein P2 (32 kDa). So, an X-ray crystallography study of isolated P2 was undertaken [57]. The failure of MR using supposedly related X-ray structures as searching models, and the difficulty of getting heavy-atom soaked crystals or suitable selenomethionine (SeMet)-derivatized X-ray data (see Sect. 7.4) required the use of the low-resolution electron density corresponding to a P2-trimer extracted from the X-ray averaged map of PM2 as searching model. The location of the two P2-trimers within the asymmetric unit of the cell by molecular replacement led to the determination of the initial phases at 7.6-Å resolution then extended to 2.5-Å by means of non-crystallographic symmetry and solvent flattening [56].

In conclusion, although the examples mentioned above do not yet represent routine procedures, the experience gained in virus phasing provides an additional structure determination tool for protein crystallography.

7.4 Dissecting Virus Structures by Combining Biochemical, Genetic and Biophysical Tools

Virus structures can be tackled by complementing not only the structural techniques mentioned above but also methods that involve the biochemical and genetic manipulation of the virion. It is clear that the prerequisite for high-quality structural data is the quality of the original biological material. The traditional ways to purify viruses are based on their sedimentation behaviours (rate zonal centrifugation) or their particle densities (equilibrium centrifugation). These can then be combined with differential centrifugation, concentration methods (*e.g.* filter devices) or precipitation (*e.g.* using different polyethylene glycols). The addition of biochemical and genetic modifications makes it possible to produce and to study different kinds of virus material from wild type and mutant virus infections, perform quantitative virus particle dissociations experiments, or use recombinant protein expression

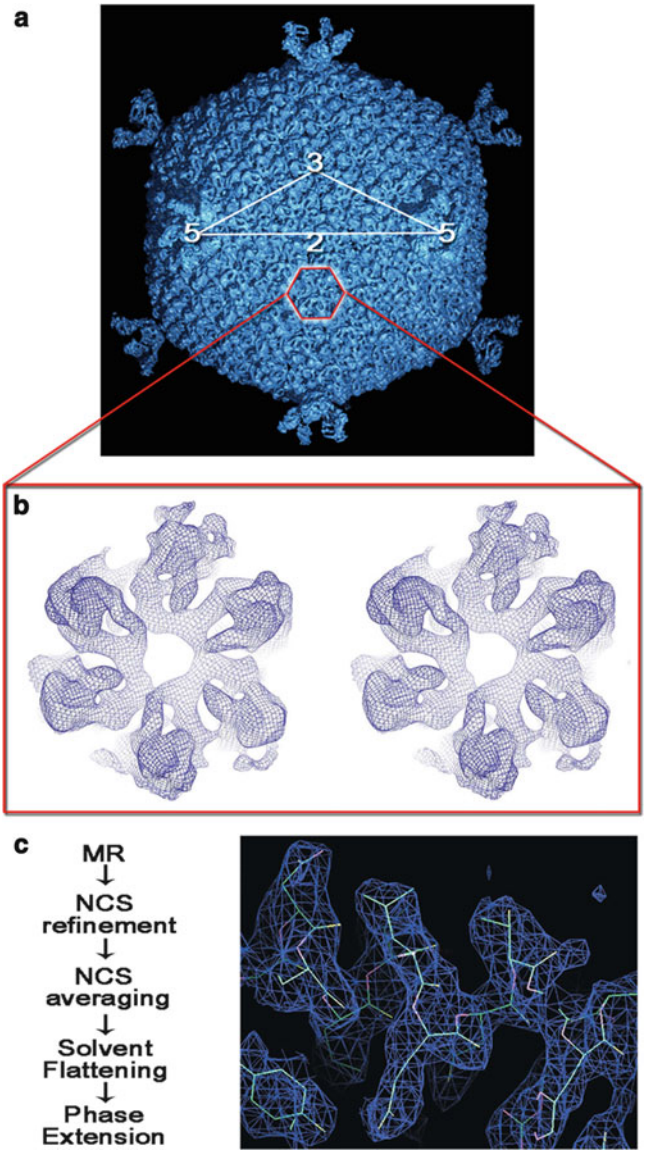


Fig. 7.5 Phasing with low-resolution phases. (**a**) X-ray-derived map of the lipid-containing bacteriophage PM2 [11] (~600-Å in diameter), which was phased using the corresponding EM map [55]; icosahedral symmetry axes in the capsid are labelled; a trimeric capsomer of the major capsid protein P2 is outlined in *red*. (**b**) Enlarged stereo-view of the extracted electron density at 7-Å corresponding to the P2 trimer used in the MR protocol. (**c**) *Left*, workflow for phasing X-ray data using the low-resolution density map in (**b**) as a searching model; *right*, experimental electron density map at 2.5-Å for the P2 major capsid protein resulting from MR, followed by phase refinement, extension and solvent flattening (Reproduced from [56]. With permission)

systems. In the following sections we will illustrate some examples where the tailored use of biochemical, genetic and biophysical techniques has provided insightful views into virus structure.

7.4.1 Classical Antibody Labeling Methods

Immuno-EM can be used as a powerful tool to localize specific proteins also asymmetrically distributed in the virion. The method relies on the specific recognition, by primary antibody, of a target protein (antigen) exposed on the virion surface. A secondary molecule recognizing the primary antibody is then added, for instance, gold-conjugated protein A, which binds with high affinity to rabbit-IgG molecules. The gold particles produce high contrast under TEM (see Sect. 7.2.1) allowing the localization of specific proteins in the virus structure, but detailed structural information is lost. Using secondary binders conjugated to gold particles of different sizes, different antigens can be co-localized within the same sample. Prior to the labelling, the virus particles are usually fixed with cross linkers, glutaraldehyde being one of the most popular choices. Recently the toolbox of cross-linkers and chemical probing protocols available has been expanded [58] allowing the exploration of novel avenues in virus biochemical manipulation. Furthermore, virus particles can be loosened *e.g.* by mild treatment with non-ionic detergent such as Triton X-100 during the sample-grid preparation to increase the accessibility of the antigens to the primary antibodies; see for example [59].

Typical problems of immune-EM, such as unspecific labeling or background can be overcome by having appropriate controls and adjusting the amount of primary antibody. The most reliable conclusions, however, can be made when immunogold labeling EM data is combined with other approaches, as demonstrated for the dsDNA bacteriophage PRD1. Combination of immunolabeling [59] with the work done with PRD1 mutants [60] verified the presence of the unique vertex structure used for genome packaging.

7.4.2 Mutagenesis Studies

Classical mutagenesis using shotgun mutagens such as ultraviolet light (UV) or N-methyl-N-nitro-N- nitrosoguanidine (NTG) and, especially, site-directed mutagenesis are powerful techniques to create virus mutants. To identify proteins and investigate their functions specific virus mutants can be created either by generating a stop codon inside the target gene (nonsense mutants) or by deleting the entire protein coding region (deletion mutants). If the mutations are lethal, rescue systems are needed in a form of suppression or complementation, which limit the usage of virus mutants to those viruses infecting a host for which a genetic system has been

established. Genetic engineering of obscure micro-organisms isolated from the environment is demanding [61]. Bacteriophage PRD1 infects gram–negative bacteria such as *Escherichia coli* and *Salmonella enterica* enabling the use of well-established genetic tools. PRD1 is one example for which numerous suppressor-sensitive mutants have been obtained and have proved invaluable in assigning functions to the corresponding viral proteins. Furthermore, the genetic removal of the PRD1 flexible receptor-binding proteins from the virion vertices was crucial in obtaining diffracting crystals leading to the determination of the X-ray virus structure at 4-Å resolution [17, 18].

Moreover mutagenesis has also allowed the insertion of green-fluorescence protein (GFP) into large and complex viruses such as herpes or vaccinia viruses – providing an important tool for studying cell interactions or virus localization within the cell (see Chap. 14). We expect that careful selection of multiple tags will also provide structural information, *e.g.* from fluorescence resonance energy transfer (FRET) and correlation spectroscopy methods. An early example comes from work on influenza virus polymerase revealing interactions in the cell [62].

Fine structure-function relationships in virus particles at the amino acid residue level can be carried out by individually replacing specific residues based on high-resolution structural information. Classically residues are replaced by alanine, removing the side chain (beyond the Cβ atom) and the intra- or intermolecular interactions it establishes with other residues, without introducing additional groups and non-native interactions, and with the minimum probability of altering the main-chain conformation of the viral proteins. For example, systematic alanine-scanning mutational analyses on the structural and functional relevance of protein-protein and protein-nucleic acid interactions in a virus particle have been carried out for FMDV and the parvovirus minute virus of mice (see Chaps. 1 and 7). Also, resolution of the crystal structure of some designed single-amino acid mutants is providing high-resolution insights into the structural bases of physical properties, assembly, conformational dynamics, and disassembly of virus particles, for example in rhinovirus and poliovirus.

7.4.3 Quantitative Dissociation Studies

The controlled removal of specific proteins from virus particles is very valuable for structural and structure-function studies (Fig. 7.6a). Using this biochemical quantitative dissociation approach several viral proteins have been assigned to their right positions in the virion. The interactions between the capsid proteins and the underlying membrane component can be disrupted by *e.g.* urea or guanidine hydrochloride treatments, as exemplified by studies on bacteriophages PM2 and PRD1, respectively, which revealed the virus membrane associated proteins and released the major capsid proteins [64, 65]. The released proteins can be further purified by standard biochemical methods for protein crystallography [37, 57, 65]. Quantitative protease treatments of the virus particles might identify structures

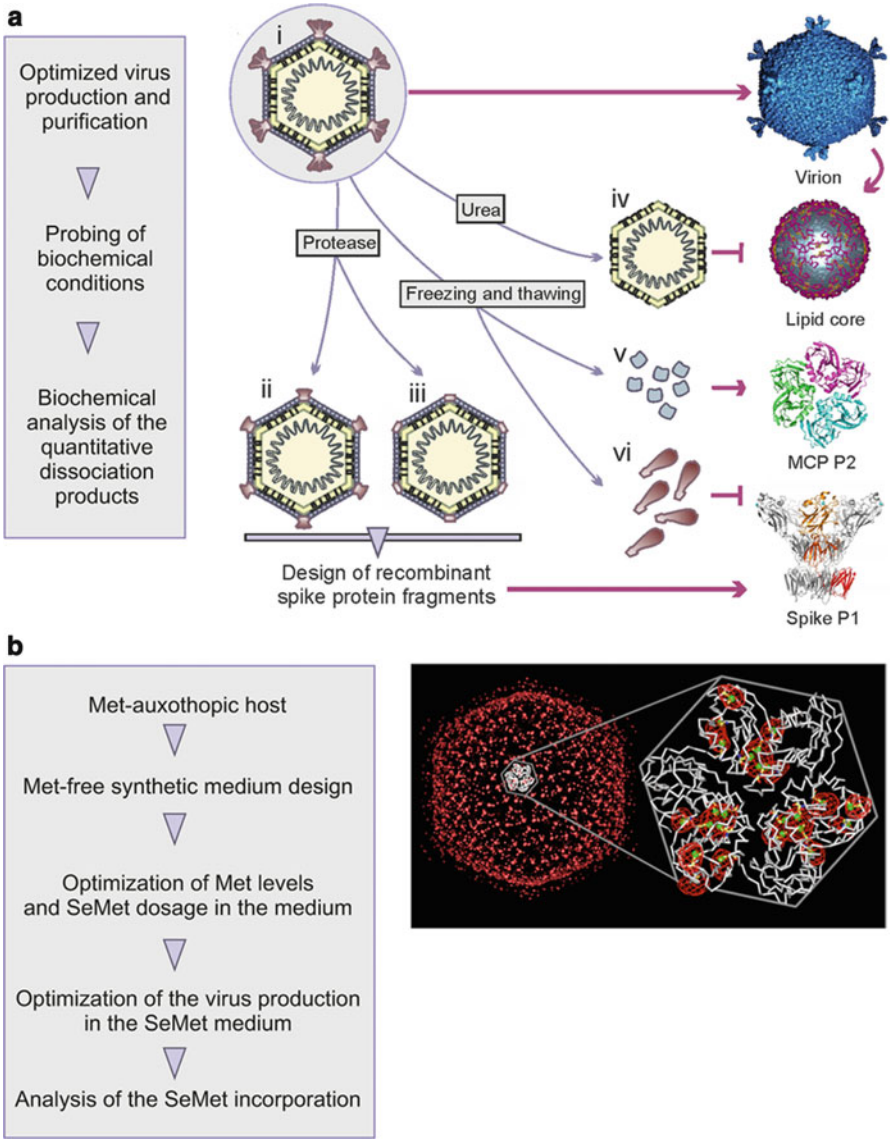


Fig. 7.6 Schematic workflow for biochemical probing. (**a**) *Left*, workflow in a study on the quantitative dissociation of virus particles. *Right*, schematic representation of the controlled dissociation of bacteriophage PM2 [63] used to solve the structures of virion components. (i) Highly purified and infectious virions used to solve the virion structure by cryo-EM [55] and X-ray crystallography [11]; (ii) proteinase K-treated virus particles devoid of the distal receptor binding domain of the spike protein P1; (iii) bromelain-treated virus particles devoid of the two outermost domains of the spike protein P1; (iv) urea treatment expose the internal lipid core particle (membrane proteins, lipids and the genome); (v) and (vi) freezing and thawing of the virions release (v) the trimeric major capsid protein (MCP), and (vi) the monomeric spike proteins.

protruding from the virion surface [55, 65]. For the analysis of membrane-containing viruses, membrane extraction with different detergent treatments can be used to selectively remove different membrane-associated proteins from the virus particle. In case of the internal-membrane containing PRD1 virus integral and peripheral membrane proteins have been separated from one another [66].

One example of a quantitative dissociation analysis is marine icosahedral dsDNA bacteriophage PM2 (Fig. 7.6b). For PM2, several unique subviral particles have been distinguished by their sedimentation properties in a sucrose gradient [55, 65]. A combination of biochemical, cryo-electron microscopic [55] and X-ray crystallographic approaches [11, 56, 57] has visualized the entire PM2 virion structure at subnanometric resolution and led to the determination of its major structural protein structures (Fig. 7.6b). Also archaeal virus SH1 [67], archaeal pleomorphic viruses [68] and *Thermus* bacteriophage P23-77 are examples of viruses for which quantitative dissociation analysis has provided useful information and significantly helped in the structural analysis [69].

7.4.4 Selenomethionine Incorporation as a Labelling Tool

Efficient selenomethionine labelling of protein species within bacteria is nowadays a standard technique but less so for the labelling of proteins expressed in mammalian cells [70]. Even more challenging is the selenomethionine labelling within viruses that when combined with X-ray crystallography is a valuable method in the structural virology. To date, only two efficient methods for selenomethione labeling of complex viruses have been reported [63, 71] and their roles in the interpretation of the electron density maps of PRD1 and PM2 were significant [11, 17, 63]. The PM2 labeling method relied on methionine auxotrophic host (Figs. 7.6c and d; [63]), whereas successful PRD1 selenemothionine incorporation was obtained with methionine prototrophic host grown in minimal medium and pulsed with selenomethione at a late stage of infection [71]. In addition to obtaining phase information in X-ray crystallography [72], selenomethionine labeling combined with difference Fourier maps can facilitate the interpretation of protein folds and even the identification of the protein species [11]. Using selenomethionated and native virus X-ray data it was possible to assign electron density to all ordered protein components in the virion based on their amino acid sequence [11, 63].

Fig. 7.6 (continued) *Far right, top to bottom*: X-ray crystallographic structure of a virion and its lipid core particle, MCP P2 trimer and the pentameric vertex-spike complex of protein P1 (Adapted from [11]. With permission) (**b**) *Left*, workflow of the selenomethionine (SeMet) labeling of bacteriophage PM2. *Right*, SeMet difference Fourier map for the entire SeMet-labeled PM2 virion contoured at 4σ (*red* mesh) with a single trimer of the MCP P2 outlined in *white*; an enlarged view of the P2 Cα backbone is shown, with the methionine residues depicted as balls-and-sticks (the incorporated selenium atoms are shown as *green spheres*) (Reproduced from [63]. With permission)

7.4.5 Some Old and New Biophysical Techniques to Study Viruses

Together with the above techniques other biophysical methods for studying virus particles in solution, including fluorescence spectroscopy, circular dichroism (CD), hydrogen exchange or limited proteolysis followed by MS (see Chap. 6) and differential scanning calorimetry (DSC) and isothermal scanning calorimetry (ITC) may be used to study different aspects of virus particle stability and conformational dynamics, and relate these properties to virus structure. DSC and ITC respectively are useful in assessing the thermal stability and conformational transitions of virus particles and in the study of virus/cell binding, entry processes and genome release. Several studies have been carried out using DSC and ITC on viruses and phages, cf. phage λ [73]. Changes in secondary structure of the capsid proteins in viral particles have been followed by far-UV CD; intrinsic fluorescence of viral particles or extrinsic fluorescence by binding dyes to the viral nucleic acid exposed on capsid disassembly have been used to follow capsid conformational rearrangements and determine the thermal or chemical stability of viral particles against dissociation (see Chap. 6).

Recently an application of the thermofluor assay used to assess protein stability before crystallization has been adapted for high-throughput screening of thermal stability of viruses [74]. This new application uses two distinct fluorescent dyes simultaneously, one with affinity to hydrophobic residues and the other to the genome. Upon heating, hydrophobic residues becoming exposed and genome release are simultaneously monitored by the fluorescence intensity and virus stability rapidly assessed.

Finally, recent developments in MS techniques have shown that it is possible to investigate very large protein complexes such as viruses by native MS (see Chap. 6). This formidable advance [75] opens the way to the analysis of virus disassembly processes traditionally considered out-of-range for this method.

7.5 Combining Electron Microscopy and Electron Tomography

7.5.1 General Approach

The understanding of the architectural and molecular organization of enveloped and complex viruses is a current challenge. Traditional approaches such as EM and more stringently X-ray crystallography require sample homogeneity in terms of structural organization and size, a premise that fails when studying those viruses that do not possess a regular shape or composition, the so-called pleomorphic viruses. Indeed such viruses often come in several sizes and shapes and the lack, until recent years, of a structural technique that could tackle such viruses had left them at the margin of structural investigation.

The advent of ET and cryo-ET [76] (Chap. 3) has allowed these difficult viruses to be tackled [77]. With this technique (for details and some examples see Chap. 3) viruses can be studied as individual particles whilst those structural elements conserved and redundant within the virus structure can be averaged, thus improving the signal-to-noise ratio and overcoming typical problems inherent to the technique such as anisotropy and incompleteness of data due to the missing wedge [76].

Thus, if a virus possesses distinct structural levels some with no symmetry (*e.g.* envelope) and others with symmetry (*e.g.* nucleocapsid) it is in principle possible to combine the cryo-ET 3D reconstruction with the other symmetry-imposed and/or averaged structural elements. These latter elements are usually at higher resolution and can be either obtained by sub-tomogram averaging within the same tomographic data collection (see for example [39, 41] and websites http://www.dynamo-em.org; http://www.biochem.mpg.de/en/rg/foerster/Content_Software/PyTom/index.html; http://www.opic.ox.ac.uk/mediawiki/index.php/Jsubtomo) or by other structural methods such as EM or X-ray crystallography. This combination will picture the architectural complexity of the virus at different resolution levels. ET is also very powerful in studying viruses in the cellular context, in particular during viral entry and replication (see also Chaps. 3 and 14). In the following section we will give a few examples of the combination of ET and EM techniques and we will illustrate the smart use of tomography in assessing particle symmetry [78].

7.5.2 Case Studies

One of the most elegant examples of combination of EM techniques in picturing a complex virus is the cryo-ET study of herpes simplex virus (HSV) [79]. In this work the regular icosahedral nucleocapsid was separated from the enclosing pleomorphic envelope in turn decorated by the tegument and glycoproteins, both indispensable for infection. Nucleocapsids were extracted from the tomograms, 60-fold averaged (as in standard icosahedral 3D reconstruction approach) and replaced in the whole virus 3D reconstructed tomogram (Fig. 7.7a), thus revealing contacts between the nucleocapsid and the tegument not seen in individual tomographic reconstructions.

Combination of ET and EM is not limited to viruses with an intrinsic symmetry mismatch. Several situations during the infection cycle induce a symmetry loss even in highly symmetric viruses. Picornaviruses are small, icosahedral, non-enveloped viruses that have been studied for several viral processes such as receptor attachment or RNA translocation. The use of decorated liposomes with poliovirus receptor CD155 allowed the imaging of the receptor bound to poliovirus both by single-particle reconstruction and electron cryo-tomography [82].

For retroviruses, cryo-ET has provided structural information of the human and simian immunodeficiency viruses (HIV/SIV) envelope spike proteins gp41/gp120 (Env) responsible for binding to the cellular receptor. The gp120/gp41 complex is randomly distributed onto the virion surface making it difficult to study [83, 84].

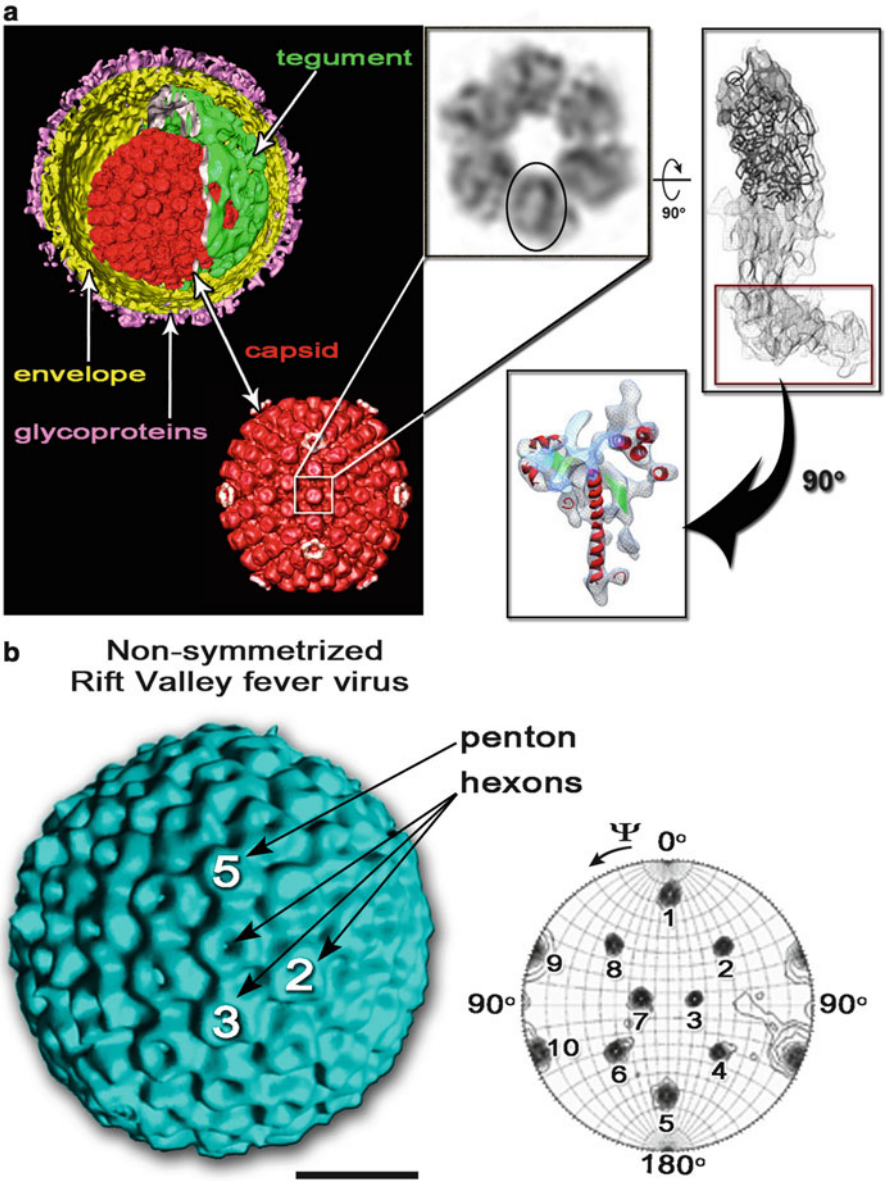


Fig. 7.7 Using X-ray tools in electron tomography. (**a**) *Top left*, herpes virus electron cryo-tomography reconstruction with the different structural elements [79]; below, herpes virus nucleocapsid reconstruction obtained by sub-tomogram averaging combined with icosahedral averaging; outlined one capsomer sitting on a 2-fold icosahedral axis (Courtesy of K. Grünewald, University of Oxford); *top center* cryo-EM derived density corresponding to this capsomer which is composed by six copies of VP5 protein; *top right*, a VP5 monomer composed of three domains (upper, middle and floor domain); the crystal structure of the upper domain is shown as a Cα ribbon model (*dark-grey*) fitted into the corresponding cryo-EM density; no X-ray structure exists for the

Subtomogram averaging of thousands of individual spikes and the imposition of the threefold symmetry gave a 3D map of sufficient quality to dock in the structure of SIV-gp120 core previously solved by X-ray crystallography [85]. Additional studies with bound receptor and antibody have elucidated key aspects of the spatial arrangement during cell and antibody recognition and conformational tuning [86]. Moreover beneath the HIV/SIV envelope, during maturation and after cleavage of the Gag polyprotein, the capsid domain forms the cone-shaped core shell organized as a hexagonal lattice. Tomographic images of this lattice followed by subtomogram averaging of the hexagonal building block allowed individual protein domains to be identified and fitted with the corresponding X-ray crystal structure [42].

Cryo-ET has also provided a 3D picture of influenza virus that displays an evident pleomorphy [87]. The two different surface proteins could be distinguished directly from the tomograms and molecular shape comparison with the crystal structures of hemagglutinin (HA) or neuraminidase (NA) allowed mapping of the individual spikes in the tomograms.

Preliminary analysis of negatively-stained samples of several bunyaviruses showed a large diversity in size, morphogenesis, structure and surface. One member of this virus family is Rift Valley fever virus (RVFV), a major pathogen of animals and humans in Africa. Cryo-ET provided the first useful 3D reconstruction of the RVFV [78], achieved by reconstructing 46 individual particles with similar overall size and shape and then averaging them together. The averaged map was Fourier transformed and the derived structure factor amplitudes used to calculate a self-rotation function [a routine procedure used in X-ray crystallography to detect the redundant presence of structural elements within the asymmetric unit of the crystal (see Chap. 4), Fig. 7.7b]. This identified five-, three- and twofold axis and allowed the unambiguous assignment of icosahedral symmetry to this virus. In turn, imposition of the 60-fold icosahedral symmetry further improved the resolution of the 3D reconstruction from 75-Å to 61-Å, revealing the capsomers arrangement and the symmetry of the lattice (triangulation number *pseudo*-T = 12). This laid the basis for structural studies at higher resolution that was achieved by the systematic single-particle sub-tomogram averaging of 267 particles, leading to a virus map at 20-Å resolution. This dramatic improvement allowed the analysis of the structure of the heterodimer formed by the G_N and G_C glycoproteins which decorate the envelope [88].

Fig. 7.7 (continued) middle domain; the floor domain is delineated by a *red rectangle* (Figure adapted from [80]. With permission). *Bottom right*, cryo-EM density corresponding to the floor domain; secondary structure templates have been fitted to suggest structural homology with the gp5 capsid protein of bacteriophage HK97 (Figure adapted from [81]. With permission). (**b**) *Left*, electron density reconstruction of RVFV by single particle electron tomography [78]; capsomers and icosahedral 2-, 3- and 5-fold symmetry axes are marked; *right*, stereographic projection of the Kappa = 120° section for the self-rotation function of the tomographic reconstruction, revealing the presence of 10 3-fold axes, and thus demonstrating the icosahedral nature of the RVFV; scale bar represents 25 nm (original images courtesy of SJ Watowich, University of Texas Medical Branch)

7.6 From Viruses in Solution to Virus in Cells

7.6.1 Virus Structures in the Cellular Context

Viruses have been traditionally observed within infected cells using either fluorescent labelled viruses and light-microscopy (resolution worse than 200-nm) or using cell-sectioning techniques for subsequent imaging in an electron microscope [4] (see Chap. 14). In this latter case different cell and sample preparation techniques have been put in place and optimized to preserve close-to-native conditions of the virus and cell. These efforts have been coupled to advances that have made the acquisition of tomographic data routine. Tomography thus allows adding depth information at higher resolution than the serial addition of ultrathin 2D sections (for details and examples see Chaps. 3 and 14).

One of the most recent additions to the available tools to investigate virus structures has been provided by synchrotron facilities through the generation of soft X-rays in the water window (wavelength 2.3–4.4 nm) that due the differential absorption of oxygen and carbon atoms (ten times lower for O than C) provide contrast for cellular tomography imaging. Virus X-ray tomography has seen pioneering applications using as model systems the virus life cycle of vaccinia virus [89] and herpesvirus [6]. These studies demonstrate that soft X-rays can provide valuable 3D structural insights into virus maturation events at the sub-nanometer level (50–30 nm resolution) for reasonably thick (5–8 μm) cells and compartments, including the cell nucleus.

Interestingly, virus crystallography is also aiming towards *in vitro* structural analysis. Indeed crystals of large viruses are extremely fragile and there are attempts to grow intracellular crystals for structure determination – a grand challenge for crystallography (see also Chap. 4). Inspirational of this new crystallogenesis strategy has been the observation that not only does protein crystallization occur *in vivo* [90, 91] but also paracrystalline arrays of virus particles often form within infected cells [4, 92]. However further developments in synchrotron radiation technology in particular in delivering high-brilliance synchrotron light, are needed to make X-ray data collection possible from these *in vivo* crystals and hopes are raised by the preliminary tests of the X-ray free electron laser (XFEL) on viruses [93, 94], that although technically successful provided very low resolution information (<200-Å) due to the limited number of incident photons in each short pulse of X-rays.

7.6.2 Methods for Cellular Landscape Scanning

One of the major difficulties in the structural interpretation of virus-infected cells is the confusion deriving from macromolecular crowding. In order to be able to provide an atlas of the different steps of a viral infection, following how the viruses enter the cell, replicate and assemble, and finally leave the cell, different timeline

points are structurally investigated during infection (see Sect. 7.7.1). For each timepoint tomograms are collected and combined computational techniques, from segmentation to template matching, are employed to detect recurrent cellular or virus structures [95, 96]. Sub-tomogram averaging of equivalent structures increases the signal-to-noise ratio and allows the improved structure to be inserted back into the cellular landscape [97, 98]. This approach can also use virus structures obtained at higher resolution by other methods such as cryo-EM or X-ray crystallography, thus providing more accurate information about the geometrical arrangement and spacing between virus particles and cellular structures. This hybrid approach fills-in information inevitably missing from grid-based ET methods (see Chaps. 3 and 14) and bridges the resolution gap between 3D cellular tomography and other methods such as single particle EM and X-ray crystallography [92].

7.7 Emerging Hybrid Methods

7.7.1 Correlative Microscopy

The analysis of biological functions from whole organisms to cells and subsequently to single molecules has been a dream for structural scientists for decades. Structural biology and virology has recently become increasingly interested in dynamic cellular and virus-cell interaction processes This requires a time-resolved approach to the visualization at the different size levels and the correlation in time and space between the different methods remains a challenge.

Correlative microscopy allows the integration of data across the microns to nanometers range, using fluorescence light microscopy and electron microscopy [6, 99, 100] (see Chap. 14). However, from the point of view of combined structural strategies it is important to underline that the aim is to identify cellular regions of interests by the use of specific chromophore labels in the light microscope and track biological events before stopping the clock and imaging the same region at much higher-resolution using an electron microscope. In this way virus-cell interactions are reconstructed in 3D and framed in time. A critical step is finding the same cellular region once the cell is within the electron microscope. To this end a common reference system between the light- and electron-microscopes must exist. Technological advances in robotics and microscopy cryo-holders have made possible data collection and a general work-flow for the procedure, pioneered by Plitzko [99] (Fig. 7.8). While correlative microscopy for 3D tomographic reconstruction is still in its infancy the potential benefits are enormous in studying virus-infected cells. Recent important contributions come, for example from the study of retroviruses where it has been possible to identify HIV particles attached to cell membranes [101] and to show the release of mutant HIV cores into the cytoplasm of host cells [102].

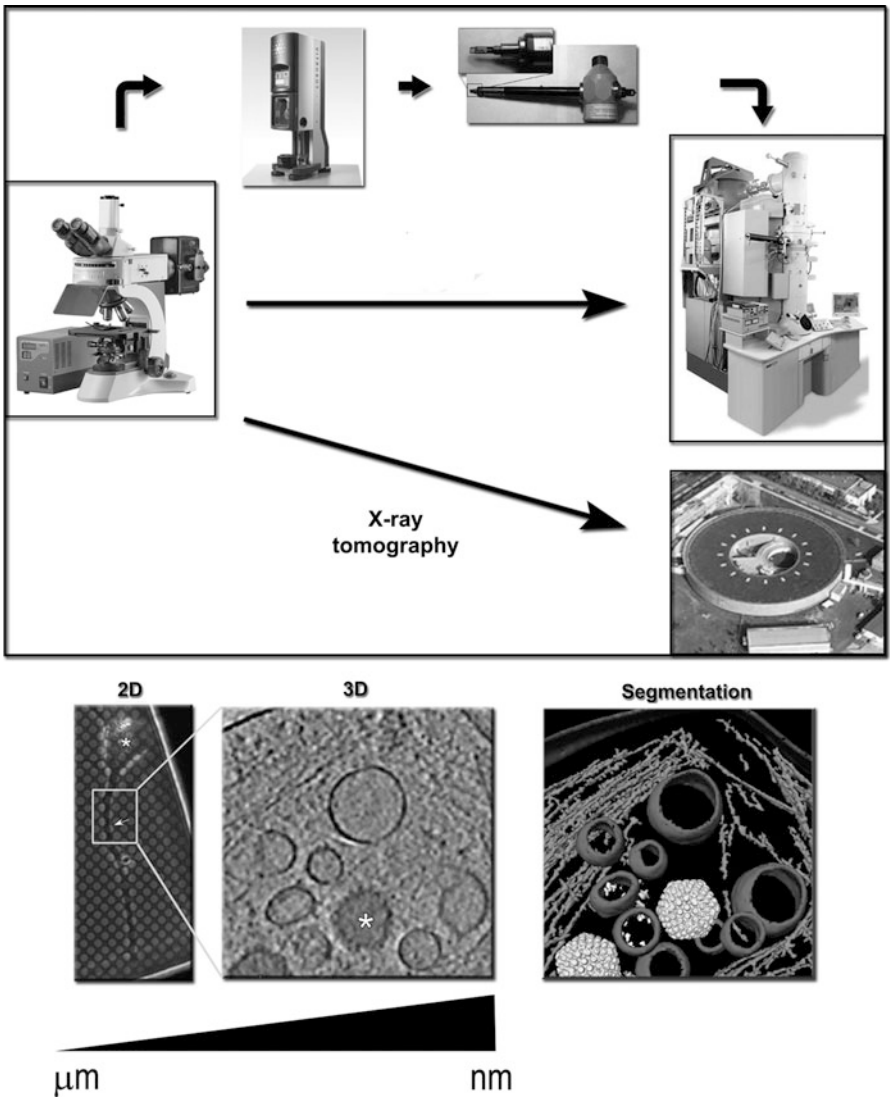


Fig. 7.8 Workflow in correlative microscopy. *Top*, Schematic of the methodological steps involved in performing correlative light and electron microscopy (CLEM). Cellular structures are sequentially identified with different microscopic techniques with increasing resolving power. Targeted processes in cells are imaged using a fluorescence microscope (*centre left*); then, the cells supported on a grid, are vitrified using a plunging freezing device (*top left*) and the grid transferred using a suitable cryo-holder (*top right*) in a field-emission-gun transmission electron microscope (*centre right*) for visualization of sub-cellular features by electron tomography; nowadays also X-ray tomography on entire cells can be performed at synchrotron facilities. *Bottom*, *left to right*: detection of the area of interest in 2D (using light microscopy; micrometer scale); visualization and 3D reconstruction by ET techniques (nanometer scale); segmentation of the different structural elements

The advent of correlative microscopy will benefit the study of those viruses that are characterized by low replication and infection efficiencies and have therefore escaped direct visualization. A clear example of this is hepatitis C virus, whose low-level replication *in vitro* has hampered the visualization of viral morphological processes by traditional ET technique.

7.7.2 Developing 3D Imaging Techniques

One of the most recent (and contentious) method developments in 3D image reconstruction algorithms that could, in principle, be used for initial low-resolution phasing is represented by ankylography [103]. Briefly, it consists in collecting on a curved detector the diffraction pattern generated by the scattering of a coherent soft X-ray laser hitting the object of interest (not necessarily crystalline). The proposed advantage of ankylography over XFEL crystallography or X-ray tomography is that, unlike these techniques, it uses a 2D spherical diffraction pattern to reconstruct the full 3D structure of the object. A proof-of-principle of this reconstruction method was the determination of poliovirus at 20–30 Å resolution [103]. So, while it might appear that ankylography might have an important impact, there is a fundamental problem in that capturing a thin spherical shell of data in reciprocal space is ultimately a very poor approximation to measuring the full 3D volume needed to produce a high resolution reconstruction [104].

As a variant on coherent X-ray scattering, ptychography, an imaging technique that was proposed in the late ‘60s by Hoppe [105] and aiming to solve the phase-problem by using an interference phenomenon has been recently re-emerging as an EM technique that could provide potentially superresolution (~5-Å) [106]. The principles behind ptychography have been also employed in X-ray tomography in life science experiments [107]. It will be interesting to see the ultimate scope of this method, however in its present form it is slow (even at a synchrotron) and requires a high radiation dose to achieve a 3D reconstruction at modest resolution.

7.8 The Biology Behind the Combined Methods

7.8.1 Implications for Drug-Design and Vaccine Development

Current methods for drug-discovery rely mainly on: (i) ligand-based design and (ii) structure-based design. In the first case the search for new compounds against a target molecule is based on previously known drugs; in the second case the search is based on the availability of the 3D structure of the target molecule (see Chap. 20 for a comprehensive view on the translational applications of structural virology

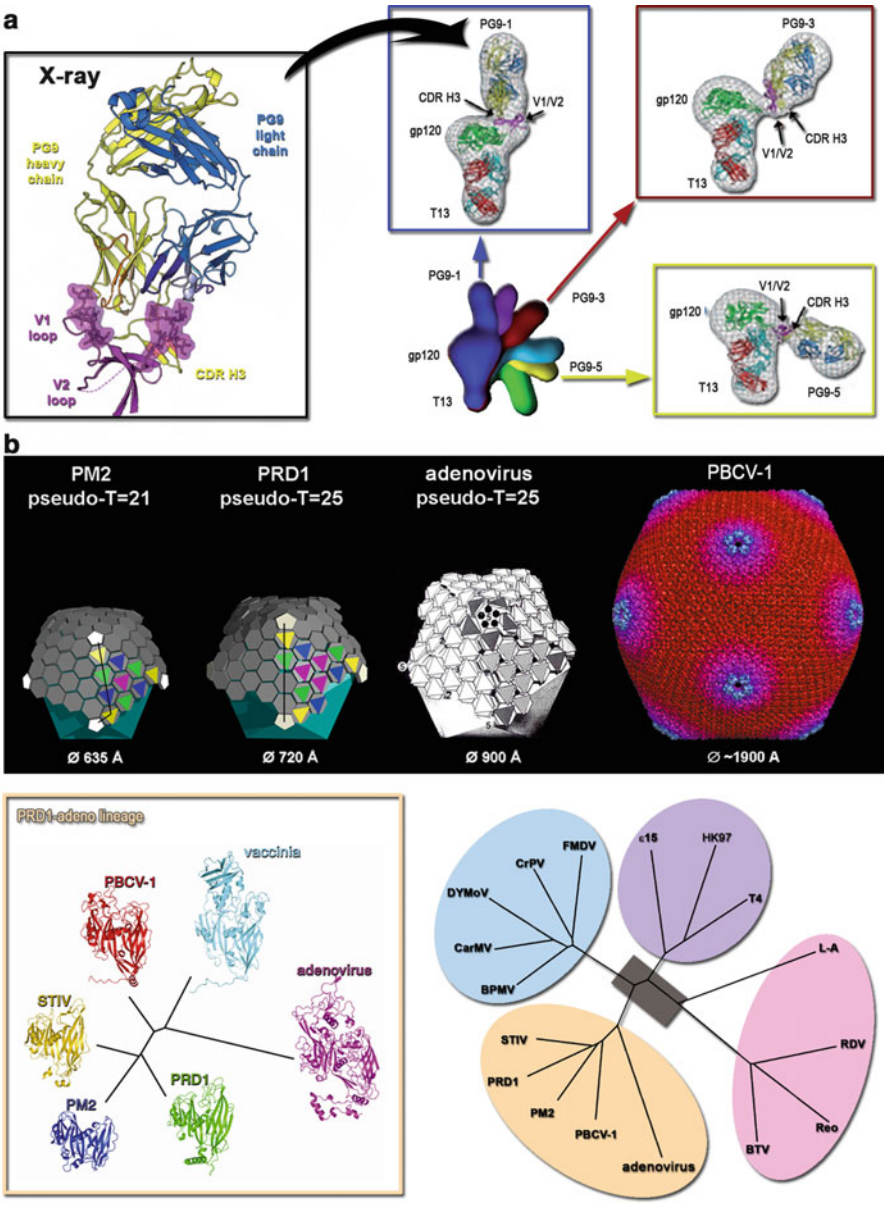


Fig. 7.9 Impact of the combination of structural methods. (**a**) *Left*, X-ray structure of the HIV-1 gp120 V1/V2 domain (in *magenta*) complexed with neutralizing antibody PG9 (*yellow*, heavy chain; *blue*, light-chain) represented as ribbons; *right*, negative stain EM reconstruction classes (six in total, colour-coded from *blue* to *green*) of ternary complexes composed by gp120, PG9 and CD4-binding-site antibody T13 with the corresponding fitting of the gp120 V1/V2-PG9 crystal structure in three of them (insets) (Adapted from [109]. With permission). (**b**) *Top left*, schematic representation of the architecture of bacteriophages PM2 and PRD1; *grey* hexagons represent the pseudo-hexameric morphology displayed by viral major capsid proteins forming trimers; the trimers composing the icosahedral asymmetric unit are represented as *colour-coded triangles*;

in antiviral research). Both screening strategies are computationally intensive. Structure-based design uses atomic or quasi atomic models to rationalize the interaction of the potential candidate drug with the target biomolecule, for example to block the binding of the virus to the cellular receptor. One of the first examples where the interdisciplinary approach of virus EM with X-ray crystallography has provided clues for the development of antiviral drugs is represented by the docking of ICAM I to human rhinovirus [26]. Indeed this quasi-atomic model (Fig. 7.3b) contributed to the development of pleconaril, a drug that inhibited the uncoating of rhinovirus upon infection by binding to the major capsid protein VP1 [108]. Although this drug was never licensed, failing in phase-clinic III studies [108], the whole process reinforced the utility of the combined approach in virus structure determination.

Undoubtedly, the interaction of viruses with their cellular receptors is one of the major structural targets in antiviral research facilitating the rational design of novel drugs aiming to halt cell entry. Similarly, studies aimed at understanding the mechanism of viral evasion of antibody neutralization, for example for HIV-1 gp120 [109], have made use of crystal structures and negative stain electron microscopy (Fig. 7.9a). Nevertheless there are other steps during virus morphogenesis that can be tackled by combining EM information and X-ray crystallography, for example targeting the ribonucleoprotein complex of influenza virus responsible for viral RNA synthesis [111]. Furthermore the structural elucidation of virus-like-particles (VLPs) as transporters of foreign antigenic epitopes (see Chap. 21) by combining cryo-EM and crystal structures of the component proteins allow the localization and investigation of the modifications introduced in the virus capsid structure. The knowledge of the conformation of the inserted epitope in the VLPs and their structural relationship with the rest of the particles directs possible improvements of construct design; examples include TBSV and Norovirus which have been used as a platform for polyvalent display of antigenic epitopes and vaccine design. Indeed structural virology has a huge potential for the structure-based development of next generation vaccines.

7.8.2 *Novel Concepts in Virus Evolution*

In the last decade the accumulation of virus and viral protein structures determined mainly by X-ray crystallography and cryo-EM or the combination of these two

Fig. 7.9 (continued) black-lines on the capsomers indicates the T numbers). *Top right*, adenovirus (Adapted from [3]), and the cryo-EM derived density of Paramecium bursaria chlorella virus-1 (PBCV-1) (image taken from http://viperdb.scripps.edu/info_page.php?VDB=1m4x). All four represented viruses are members of the PRD1-adeno viral lineage. *Bottom left*, structure-based phylogenetic tree of viral capsid protein members of the PRD1-adeno viral lineage (Reproduced from [110]. With permission). *Bottom right*, the four viral lineages so far grouped using the phylogenetic structural approach; each lineage includes some virus/phage representatives (PRD1-adeno viral lineage, *ocre*; BTV-like lineage, *pink*; Picorna-like lineage, *light-blue*; HK97-lineage, *violet*) (Reproduced with permission from Abrescia et al. 2010)

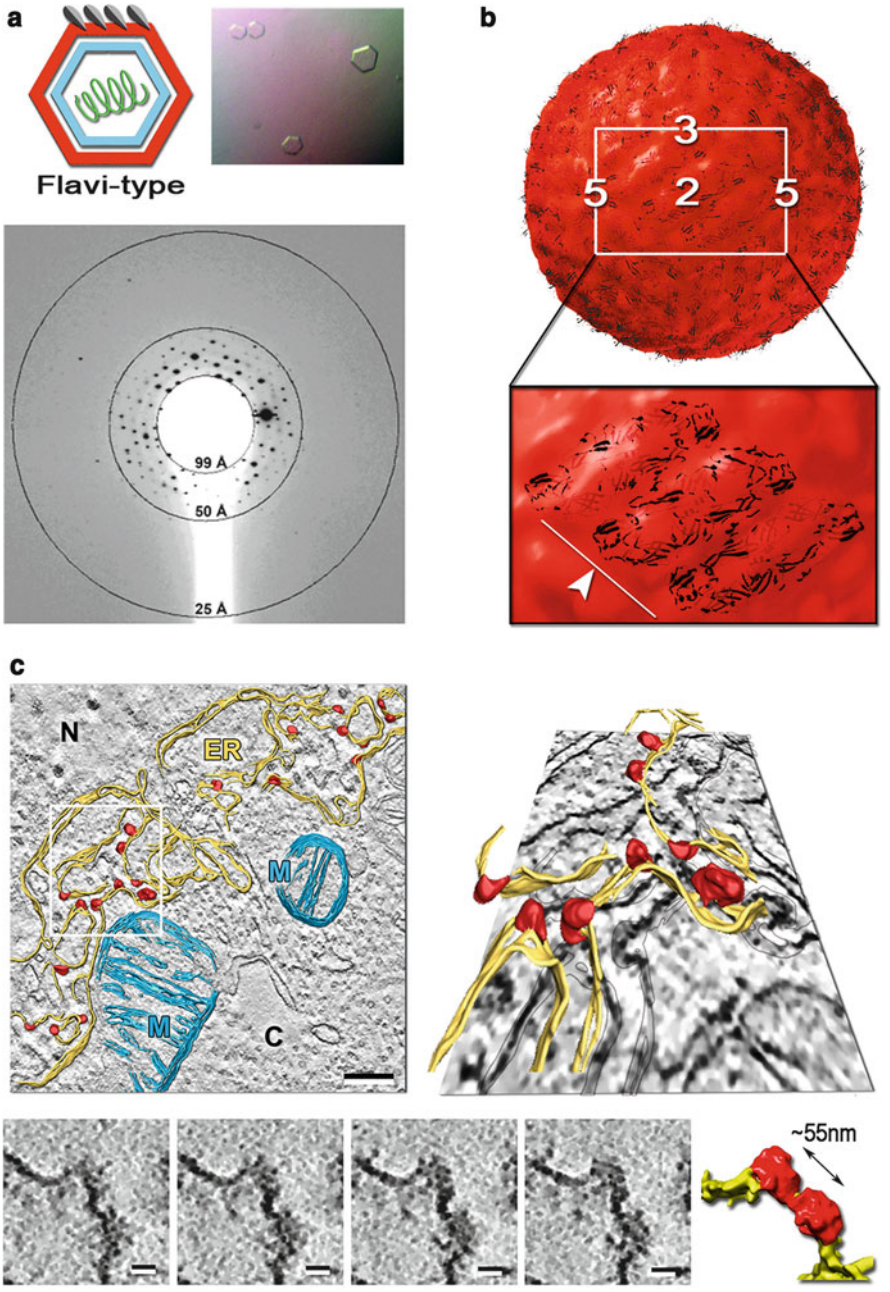


Fig. 7.10 Some of the challenges ahead. (**a**) *Top left*, scheme of a flavivirus; the envelope is shown in *red*, the glycoproteins in *grey* or *light-grey*, the capsid in *cyan* and the genomic RNA in *green*. *Top right*, one of the different morphologies of WNV crystals. *Bottom*, diffraction pattern at ~25-Å resolution corresponding to the crystals shown on top (Adapted from [113]. With permission). (**b**) *Top*, three-dimensional cryo-EM reconstruction of HCV-like-particles (HCV-LPs)

techniques has allowed us to look at the virus world (virosphere) on the basis of their structural relationships. This has led to the emergence of a novel concept in virus evolution that uses specific structural markers, the so-called virus-self elements (for example the fold of the capsid protein) to infer common ancestry [38]. Although the observation that viruses infecting different species (cf. SBMV vs. rhinovirus) displayed a similar structural fold in their capsid protein (single jelly-roll) goes back to 1985 [16], it was only in 2002 [38] that the capsid-centered view in virus classification was formalized as a precise hypothesis and through the years validated by the systematic experimental approach of structure comparisons [112]. This novel taxonomic approach allows us to simplify the organization of the virosphere and pinpoint relationships between viruses that infect different hosts across the three domains of life. A clear example of this is represented by adenovirus and PRD1, which infect vertebrate and bacteria, respectively. Their capsid proteins share a double jelly-roll fold and the capsids share common architectural principles [17] (Fig. 7.9b). The power of this structural classification is apparent in the fact that almost half of the virus families currently described, many of which do not belong to any order according to the International Committee on Taxonomy of Viruses (ICTV) classification scheme (see Chap. 1), can be traced back to one of the four viral lineages so far described: (*i*) PRD1/adenovirus-like, (*ii*) Blue-tongue-virus-like, (*iii*) HK97-like and (*iv*) Picorna-like [112] (Fig. 7.9b). It is unclear how many more virus lineages will be unveiled by the structure-based classification method since this depends on the discovery of new viral protein folds. Finally, this structure-based taxonomy of viruses opens the way to conceptualize anti-viral strategies that bridge viruses that were considered unrelated until now.

7.8.3 Perspectives and Challenges Ahead

Much progress has been made in the understanding virus biology thanks to the combination of biochemical, genetic and structural techniques. EM and X-ray crystallography have played (and still do) a dominant role in describing and

Fig. 7.10 (continued) (~500-Å in diameter) at 30-Å resolution; the icosahedral symmetry axes are indicated; *Bottom*, close-up of a HCV-LP capsid region; dengue virus glycoproteins (*black* cartoon models, indicated with a *white arrow-head* and *line*) have been fitted within the HCV-LP cryo-EM density (Adapted from [115]. With permission). (**c**) *Top*, electron tomographic reconstruction of Sf9 cells showing the cellular ER landscape when expressing hepatitis C virus-like-particles (HCV-LPs), and segmentation of its major structural elements. *Top left*, slice through the center of a denoised tomogram with superimposed rendering of the convoluted ER membranes in *gold*, budding HCV-LPs in *red* and mitochondria in *slate-blue*. Scale-bar = 200 nm. *Top right*, perspective view of the region outlined by a white rectangle. N = nucleus, C = cytoplasm, M = mitochondria, ER = endoplasmic reticulum. *Bottom*, four consecutive tomographic slices showing coalescence of HCV-LPs buds together with their isosurface representation and segmentation on the right. Scale-bar = 20 nm (Reproduced from [116]. With permission)

providing high-resolution viral structures, whilst their hybridization has provided insights into the mechanisms of virus-host recognition, antibody neutralization, antigen display, virus assembly and maturation and other properties and functions of virus particles (see Part III of this book, Chaps. 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19). Nevertheless, efforts in the past have focused mainly on virus particles that display icosahedral morphology (about half the virus families known), which are easier to analyse, as described above.

Structural virology has still several challenges ahead, including the structural characterization at atomic resolution of enveloped viruses, such as members of the *Flaviviridae*, *Alphaviridae* and *Bunyaviridae* families, and even more complex viruses, such as the poxviruses.

Attempts to crystallize WNV, a member of the *Flaviviridae* family, have been carried out [113] and indeed crystals were obtained and diffracted to 25-Å resolution – not competitive with the 14-Å resolution achieved by cryo-EM [114] (Fig. 7.10a). Another structural challenge is the biomedically very important hepatitis C virus (HCV) whose architecture, structure of its glycoproteins and assembly are not understood. A wealth of biochemical information has been gathered on the HCV life-cycle in past and recent years but only very recently, have efforts focused on the structural characterization of the virus [115–117] (Fig. 7.10b).

Technological developments with synchrotron radiation with the advent of XFEL and the *in-situ* diffraction [118], or the routine use in EM of the phase-plate contrast technology [119], the new direct detection cameras and correlative microscopy will undoubtedly shape our investigation on virus structures in the near future.

To sum up, the hybridization of approaches in structural virology is a powerful method in generating knowledge and tools serving in the fight against viral infections. These contributions are fundamental to the health of humans and animals, and will help to create a "Systems Virology" vision capturing the complex biology of viruses.

Acknowledgements DB acknowledges support from the CIC bioGUNE and HMO thanks the University of Helsinki for funding to EU ESFRI Instruct Centre for Virus Production and Purification. DIS is supported by the UK MRC. This work was enabled by the Spanish Ministerio de Ciencia y Innovacion (BFU2009-08123), Spanish Ministerio de Economia y Competitividad (BFU2012-33947) and the Basque Government (PI2010-20) grants to NGAA.

References and Further Reading

1. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G (1978) Tomato bushy stunt virus at 2.9 Å resolution. Nature 276:368–373
2. Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AG, Rayment I, Rossmann MG, Suck D, Tsukihara T (1980) Structure of southern bean mosaic virus at 2.8Å resolution. Nature 286:33–39
3. Roberts MM, White JL, Grutter MG, Burnett RM (1986) Three-dimensional structure of the adenovirus major coat protein hexon. Science 232:1148–1151

4. Goldsmith CS, Miller SE (2009) Modern uses of electron microscopy for detection of viruses. Clin Microbiol Rev 22:552–563
5. Jiang W, Baker ML, Jakana J, Weigele PR, King J, Chiu W (2008) Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. Nature 451:1130–1134
6. Hagen C, Guttmann P, Klupp B, Werner S, Rehbein S, Mettenleiter TC, Schneider G, Grünewald K (2012) Correlative VIS-fluorescence and soft X-ray cryo-microscopy/tomography of adherent cells. J Struct Biol 177:193–201
7. Svergun DI, Koch MH (2003) Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys 66:1735–1782
8. Svergun DI, Koch MH (2002) Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol 12:654–660
9. Stubbs G (2001) Fibre diffraction studies of filamentous viruses. Rep Prog Phys 64:1389–1425
10. Harrison SC, Caspar DL, Camerini-Otero RD, Franklin RM (1971) Lipid and protein arrangement in bacteriophage PM2. Nat New Biol 229:197–201
11. Abrescia NG, Grimes JM, Kivela HM, Assenberg R, Sutton GC, Butcher SJ, Bamford JK, Bamford DH, Stuart DI (2008) Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. Mol Cell 31:749–761
12. Canady MA, Tsuruta H, Johnson JE (2001) Analysis of rapid, large-scale protein quaternary structural changes: time-resolved X-ray solution scattering of Nudaurelia capensis omega virus (NomegaV) maturation. J Mol Biol 311:803–814
13. Fry EE, Abrescia NGA, Stuart DI (2007) Virus crystallography. In: Sanderson M, Skelly J (eds) Macromolecular crystallography: conventional and high-throughput methods – a practical approach. Oxford University Press, Oxford
14. Tuthill TJ, Harlos K, Walter TS, Knowles NJ, Groppelli E, Rowlands DJ, Stuart DI, Fry EE (2009) Equine rhinitis A virus and its low pH empty particle: clues towards an aphthovirus entry mechanism? PLoS Pathog 5:e1000620
15. Hogle JM, Chow M, Filman DJ (1985) Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358–1365
16. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht H, Johnson JJ, Kamer G, Luo M, Mosser AG, Rueckert RR, Sherry B, Vriend G (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317:145–153
17. Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martín C, Burnett RM, Stuart DI, Bamford DH, Bamford JK (2004) Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432:68–74
18. Cockburn JJ, Abrescia NG, Grimes JM, Sutton GC, Diprose JM, Benevides JM, Thomas GJ Jr, Bamford JK, Bamford DH, Stuart DI (2004) Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature 432:122–125
19. Liu H, Jin L, Koh SB, Atanasov I, Schein S, Wu L, Zhou ZH (2010) Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science 329:1038–1043
20. Reddy VS, Natchiar SK, Stewart PL, Nemerow GR (2010) Crystal structure of human adenovirus at 3.5 A resolution. Science 329:1071–1075
21. Perrakis A, Daenke S, Stuart DI, Sussman JL (2011) From SPINE to SPINE-2 complexes and beyond. J Struct Biol 175:105
22. Axford D, Owen RL, Aishima J, Foadi J, Morgan AW, Robinson JI, Nettleship JE, Owens RJ, Moraes I, Fry EE, Grimes JM, Harlos K, Kotecha A, Ren J, Sutton G, Walter TS, Stuart DI, Evans G (2012) In situ macromolecular crystallography using microbeams. Acta Crystallogr D Biol Crystallogr 68:592–600
23. Wang X, Peng W, Ren J, Hu Z, Xu J, Lou Z, Li X, Yin W, Shen X, Porta C, Walter TS, Evans G, Axford D, Owen R, Rowlands DJ, Wang J, Stuart DI, Fry EE, Rao Z (2012) A sensor-

adaptor mechanism for enterovirus uncoating from structures of EV71. Nat Struct Mol Biol 19:424–429
24. Evans G, Axford D, Owen RL (2011) The design of macromolecular crystallography diffraction experiments. Acta Crystallogr D Biol Crystallogr 67:261–270
25. Rossmann MG (2000) Fitting atomic models into electron-microscopy maps. Acta Crystallogr D Biol Crystallogr 56:1341–1349
26. Kolatkar PR, Bella J, Olson NH, Bator CM, Baker TS, Rossmann MG (1999) Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. EMBO J 18:6249–6259
27. Gilbert RJ, Grimes JM, Stuart DI (2003) Hybrid vigor: hybrid methods in viral structure determination. Adv Protein Chem 64:37–91
28. Topf M, Sali A (2005) Combining electron microscopy and comparative protein structure modeling. Curr Opin Struct Biol 15:578–585
29. Kaufmann KW, Lemmon GH, Deluca SL, Sheehan JH, Meiler J (2010) Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry 49:2987–2998
30. Franklin RE, Harrison SC, Pettersson U, Philipson L, Branden CJ, Werner P (1971) Structural studies on the adenovirus hexon. Cold Spring Harb Symp Quant Biol 36:503–510
31. Hewat EA, Verdaguer N, Fita I, Blakemore W, Brookes S, King A, Newman J, Domingo E, Mateu MG, Stuart DI (1997) Structure of the complex of an Fab fragment of a neutralizing antibody with foot-and-mouth disease virus: positioning of a highly mobile antigenic loop. EMBO J 16:1492–1500
32. Yu IM, Zhang W, Holdaway HA, Li L, Kostyuchenko VA, Chipman PR, Kuhn RJ, Rossmann MG, Chen J (2008) Structure of the immature dengue virus at low pH primes proteolytic maturation. Science 319:1834–1837
33. Li L, Lok SM, Yu IM, Zhang Y, Kuhn RJ, Chen J, Rossmann MG (2008) The flavivirus precursor membrane-envelope protein complex: structure and maturation. Science 319:1830–1834
34. Cherrier MV, Kaufmann B, Nybakken GE, Lok SM, Warren JT, Chen BR, Nelson CA, Kostyuchenko VA, Holdaway HA, Chipman PR, Kuhn RJ, Diamond MS, Rossmann MG, Fremont DH (2009) Structural basis for the preferential recognition of immature flaviviruses by a fusion-loop antibody. EMBO J 28:3269–3276
35. Bamford DH, Caldentey J, Bamford JK (1995) Bacteriophage PRD1: a broad host range dsDNA tectivirus with an internal membrane. Adv Virus Res 45:281–319
36. San Martin C, Burnett RM, de Haas F, Heinkel R, Rutten T, Fuller SD, Butcher SJ, Bamford DH (2001) Combined EM/X-Ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. Structure 9:917–930
37. Benson SD, Bamford JK, Bamford DH, Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell 98:825–833
38. Bamford DH, Burnett RM, Stuart DI (2002) Evolution of viral structure. Theor Popul Biol 61:461–470
39. Nickell S, Forster F, Linaroudis A, Net WD, Beck F, Hegerl R, Baumeister W, Plitzko JM (2005) TOM software toolbox: acquisition and analysis for electron tomography. J Struct Biol 149:227–234
40. Scheres SH, Melero R, Valle M, Carazo JM (2009) Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. Structure 17:1563–1572
41. Castaño-Diez D, Kudryashev M, Arheit M, Stahlberg H (2012) Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. J Struct Biol 178:139–151
42. Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, Kräusslich HG (2009) Structure and assembly of immature HIV. Proc Natl Acad Sci U S A 106:11090–11095

43. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, Sali A, Rout MP (2007) The molecular architecture of the nuclear pore complex. Nature 450:695–701
44. Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355:472–475
45. Tsao J, Chapman MS, Rossmann MG (1992) Ab initio phase determination for viruses with high symmetry: a feasibility study. Acta Crystallogr A 48(Pt 3):293–301
46. Chapman MS, Tsao J, Rossmann MG (1992) Ab initio phase determination for spherical viruses: parameter determination for spherical-shell models. Acta Crystallogr A 48(Pt 3): 301–312
47. Rossmann MG (1995) Ab initio phase determination and phase extension using non-crystallographic symmetry. Curr Opin Struct Biol 5:650–655
48. Rossmann MG (1990) The molecular replacement method. Acta Crystallogr A 46(Pt 2):73–82
49. Acharya R, Fry E, Stuart D, Fox G, Rowlands D, Brown F (1989) The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. Nature 337:709–716
50. Villeret V, Tricot C, Stalon V, Dideberg O (1995) Crystal structure of Pseudomonas aeruginosa catabolic ornithine transcarbamoylase at 3.0 Å resolution: a different oligomeric organization in the transcarbamoylase family. Proc Natl Acad Sci U S A 92:10762–10766
51. Dodson EJ (2001) Using electron-microscopy images as a model for molecular replacement. Acta Crystallogr D Biol Crystallogr 57:1405–1409
52. Navaza J (2008) Combining X-ray and electron-microscopy data to solve crystal structures. Acta Crystallogr D Biol Crystallogr 64:70–75
53. Trapani S, Schoehn G, Navaza J, Abergel C (2010) Macromolecular crystal data phased by negative-stained electron-microscopy reconstructions. Acta Crystallogr D Biol Crystallogr 66:514–521
54. Hao Q, Dodd FE, Grossmann JG, Hasnain SS (1999) Ab initio phasing using molecular envelope from solution X-ray scattering. Acta Crystallogr D Biol Crystallogr 55:243–246
55. Huiskonen JT, Kivelä HM, Bamford DH, Butcher SJ (2004) The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. Nat Struct Mol Biol 11:850–856
56. Abrescia NG, Grimes JM, Oksanen HM, Bamford JK, Bamford DH, Stuart DI (2011) The use of low-resolution phasing followed by phase extension from 7.6 to 2.5 Å resolution with noncrystallographic symmetry to solve the structure of a bacteriophage capsid protein. Acta Crystallogr D Biol Crystallogr 67:228–232
57. Abrescia NG, Kivelä HM, Grimes JM, Bamford JK, Bamford DH, Stuart DI (2005) Preliminary crystallographic analysis of the major capsid protein P2 of the lipid-containing bacteriophage PM2. Acta Crystallogr Sect F Struct Biol Cryst Commun 61:762–765
58. Leitner A, Reischl R, Walzthoeni T, Herzog F, Bohn S, Förster F, Aebersold R (2012) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. Mol Cell Proteomics 11(M111):014126
59. Gowen B, Bamford JK, Bamford DH, Fuller SD (2003) The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. J Virol 77:7863–7871
60. Strömsten NJ, Bamford DH, Bamford JK (2003) The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. J Virol 77:6314–6321
61. Kivelä HM, Madonna S, Krupovic M, Tutino ML, Bamford JK (2008) Genetics for Pseudoalteromonas provides tools to manipulate marine bacterial virus PM2. J Bacteriol 190:1298–1307
62. Avilov SV, Moisy D, Munier S, Schraidt O, Naffakh N, Cusack S (2011) Replication-competent influenza A virus that encodes a split-green fluorescent protein-tagged PB2 polymerase subunit allows live-cell imaging of the virus life cycle. J Virol 86:1433–1448

63. Kivelä HM, Abrescia NG, Bamford JK, Grimes JM, Stuart DI, Bamford DH (2008) Selenomethionine labeling of large biological macromolecular complexes: probing the structure of marine bacterial virus PM2. J Struct Biol 161:204–210
64. Bamford D, Mindich L (1982) Structure of the lipid-containing bacteriophage PRD1: disruption of wild-type and nonsense mutant phage particles with guanidine hydrochloride. J Virol 44:1031–1038
65. Kivelä HM, Kalkkinen N, Bamford DH (2002) Bacteriophage PM2 has a protein capsid surrounding a spherical proteinaceous lipid core. J Virol 76:8169–8178
66. Caldentey J, Luo C, Bamford DH (1993) Dissociation of the lipid-containing bacteriophage PRD1: effects of heat, pH, and sodium dodecyl sulfate. Virology 194:557–563
67. Kivelä HM, Roine E, Kukkaro P, Laurinavicius S, Somerharju P, Bamford DH (2006) Quantitative dissociation of archaeal virus SH1 reveals distinct capsid proteins and a lipid core. Virology 356:4–11
68. Pietilä MK, Atanasova NS, Manole V, Liljeroos L, Butcher SJ, Oksanen HM, Bamford DH (2012) Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. J Virol 86:5067–5079
69. Rissanen I, Pawlowski A, Harlos K, Grimes JM, Stuart DI, Bamford JKH (2012) Crystallization and preliminary crystallographic analysis of the major capsid proteins VP16 and VP17 of bacteriophage P23-77. Acta Crystallogr F68:580–583
70. Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr 62:1243–1250
71. Benson SD, Bamford JK, Bamford DH, Burnett RM (2002) The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1.65 A resolution. Acta Crystallogr D Biol Crystallogr 58:39–59
72. Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254:51–58
73. Jeembaeva M, Jonsson B, Castelnovo M, Evilevitch A (2010) DNA heats up: energetics of genome ejection from phage revealed by isothermal titration calorimetry. J Mol Biol 395:1079–1087
74. Walter TS, Ren J, Tuthill TJ, Rowlands DJ, Stuart DI, Fry EE (2012) A plate-based high throughput assay for virus stability and vaccine formulation. J Virol Methods 185(1):166–170
75. Shoemaker GK, van Duijn E, Crawford SE, Uetrecht C, Baclayon M, Roos WH, Wuite GJ, Estes MK, Prasad BV, Heck AJ (2010) Norwalk virus assembly and stability monitored by mass spectrometry. Mol Cell Proteomics 9:1742–1751
76. Lucic V, Forster F, Baumeister W (2005) Structural studies by electron tomography: from cells to molecules. Annu Rev Biochem 74:833–865
77. Subramaniam S, Bartesaghi A, Liu J, Bennett AE, Sougrat R (2007) Electron tomography of viruses. Curr Opin Struct Biol 17:596–602
78. Freiberg AN, Sherman MB, Morais MC, Holbrook MR, Watowich SJ (2008) Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. J Virol 82:10341–10348
79. Grünewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, Steven AC (2003) Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science 302:1396–1398
80. Bowman BR, Baker ML, Rixon FJ, Chiu W, Quiocho FA (2003) Structure of the herpesvirus major capsid protein. EMBO J 22:757–765
81. Baker ML, Jiang W, Rixon FJ, Chiu W (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. J Virol 79:14967–14970
82. Bostina M, Bubeck D, Schwartz C, Nicastro D, Filman DJ, Hogle JM (2007) Single particle cryoelectron tomography characterization of the structure and structural variability of poliovirus-receptor-membrane complex at 30 A resolution. J Struct Biol 160:200–210
83. Zhu P, Liu J, Bess J Jr, Chertova E, Lifson JD, Grisé H, Ofek GA, Taylor KA, Roux KH (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature 441:847–852

84. Zanetti G, Briggs JA, Grünewald K, Sattentau QJ, Fuller SD (2006) Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. PLoS Pathog 2:e83
85. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC (2005) Structure of an unliganded simian immunodeficiency virus gp120 core. Nature 433:834–841
86. Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S (2008) Molecular architecture of native HIV-1 gp120 trimers. Nature 455:109–113
87. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci U S A 103:19123–19127
88. Huiskonen JT, Overby AK, Weber F, Grünewald K (2009) Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: evidence for GN-GC glycoprotein heterodimers. J Virol 83:3762–3769
89. Carrascosa JL, Chichon FJ, Pereiro E, Rodriguez MJ, Fernandez JJ, Esteban M, Heim S, Guttmann P, Schneider G (2009) Cryo-X-ray tomography of vaccinia virus membranes and inner compartments. J Struct Biol 168:234–239
90. Ji X, Sutton G, Evans G, Axford D, Owen R, Stuart DI (2009) How baculovirus polyhedra fit square pegs into round holes to robustly package viruses. EMBO J 29:505–514
91. Koopmann R, Cupelli K, Redecke L, Nass K, Deponte DP, White TA, Stellato F, Rehders D, Liang M, Andreasson J, Aquila A, Bajt S, Barthelmess M, Barty A, Bogan MJ, Bostedt C, Boutet S, Bozek JD, Caleman C, Coppola N, Davidsson J, Doak RB, Ekeberg T, Epp SW, Erk B, Fleckenstein H, Foucar L, Graafsma H, Gumprecht L, Hajdu J, Hampton CY, Hartmann A, Hartmann R, Hauser G, Hirsemann H, Holl P, Hunter MS, Kassemeyer S, Kirian RA, Lomb L, Maia FR, Kimmel N, Martin AV, Messerschmidt M, Reich C, Rolles D, Rudek B, Rudenko A, Schlichting I, Schulz J, Seibert MM, Shoeman RL, Sierra RG, Soltau H, Stern S, Strüder L, Timneanu N, Ullrich J, Wang X, Weidenspointner G, Weierstall U, Williams GJ, Wunderer CB, Fromme P, Spence JC, Stehle T, Chapman HN, Betzel C, Duszenko M (2012) *In vivo* protein crystallization opens new routes in structural biology. Nat Methods 9:259–262
92. Iwasaki K, Omura T (2010) Electron tomography of the supramolecular structure of virus-infected cells. Curr Opin Struct Biol 20:632–639
93. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, Hunter MS, Schulz J, DePonte DP, Weierstall U, Doak RB, Maia FR, Martin AV, Schlichting I, Lomb L, Coppola N, Shoeman RL, Epp SW, Hartmann R, Rolles D, Rudenko A, Foucar L, Kimmel N, Weidenspointner G, Holl P, Liang M, Barthelmess M, Caleman C, Boutet S, Bogan MJ, Krzywinski J, Bostedt C, Bajt S, Gumprecht L, Rudek B, Erk B, Schmidt C, Hömke A, Reich C, Pietschner D, Strüder L, Hauser G, Gorke H, Ullrich J, Herrmann S, Schaller G, Schopper F, Soltau H, Kühnel KU, Messerschmidt M, Bozek JD, Hau-Riege SP, Frank M, Hampton CY, Sierra RG, Starodub D, Williams GJ, Hajdu J, Timneanu N, Seibert MM, Andreasson J, Rocker A, Jönsson O, Svenda M, Stern S, Nass K, Andritschke R, Schröter CD, Krasniqi F, Bott M, Schmidt KE, Wang X, Grotjohann I, Holton JM, Barends TR, Neutze R, Marchesini S, Fromme R, Schorb S, Rupp D, Adolph M, Gorkhover T, Andersson I, Hirsemann H, Potdevin G, Graafsma H, Nilsson B, Spence JC (2011) Femtosecond X-ray protein nanocrystallography. Nature 470:73–77
94. Seibert MM, Ekeberg T, Maia FR, Svenda M, Andreasson J, Jönsson O, Odić D, Iwan B, Rocker A, Westphal D, Hantke M, DePonte DP, Barty A, Schulz J, Gumprecht L, Coppola N, Aquila A, Liang M, White TA, Martin A, Caleman C, Stern S, Abergel C, Seltzer V, Claverie JM, Bostedt C, Bozek JD, Boutet S, Miahnahri AA, Messerschmidt M, Krzywinski J, Williams G, Hodgson KO, Bogan MJ, Hampton CY, Sierra RG, Starodub D, Andersson I, Bajt S, Barthelmess M, Spence JC, Fromme P, Weierstall U, Kirian R, Hunter M, Doak RB, Marchesini S, Hau-Riege SP, Frank M, Shoeman RL, Lomb L, Epp SW, Hartmann R, Rolles D, Rudenko A, Schmidt C, Foucar L, Kimmel N, Holl P, Rudek B, Erk B, Hömke A, Reich C, Pietschner D, Weidenspointner G, Strüder L, Hauser G, Gorke H, Ullrich J, Schlichting I, Herrmann S, Schaller G, Schopper F, Soltau H, Kühnel KU, Andritschke R, Schröter CD,

Krasniqi F, Bott M, Schorb S, Rupp D, Adolph M, Gorkhover T, Hirsemann H, Potdevin G, Graafsma H, Nilsson B, Chapman HN, Hajdu J (2011) Single mimivirus particles intercepted and imaged with an X-ray laser. Nature 470:78–81
95. Lebbink MN, Geerts WJ, van der Krift TP, Bouwhuis M, Hertzberger LO, Verkleij AJ, Koster AJ (2007) Template matching as a tool for annotation of tomograms of stained biological structures. J Struct Biol 158:327–335
96. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC (2010) Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. J Struct Biol 170:427–438
97. Liljeroos L, Huiskonen JT, Ora A, Susi P, Butcher SJ (2011) Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions. Proc Natl Acad Sci U S A 108:18085–18090
98. Brandt F, Carlson LA, Hartl FU, Baumeister W, Grünewald K (2010) The three-dimensional organization of polyribosomes in intact human cells. Mol Cell 39:560–569
99. Plitzko JM, Rigort A, Leis A (2009) Correlative cryo-light microscopy and cryo-electron tomography: from cellular territories to molecular landscapes. Curr Opin Biotechnol 20:83–89
100. van Driel LF, Valentijn JA, Valentijn KM, Koning RI, Koster AJ (2009) Tools for correlative cryo-fluorescence microscopy and cryo-electron tomography applied to whole mitochondria in human endothelial cells. Eur J Cell Biol 88:669–684
101. Kukulski W, Schorb M, Welsch S, Picco A, Kaksonen M, Briggs JA (2011) Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. J Cell Biol 192:111–119
102. Jun S, Ke D, Debiec K, Zhao G, Meng X, Ambrose Z, Gibson GA, Watkins SC, Zhang P (2011) Direct visualization of HIV-1 with correlative live-cell microscopy and cryo-electron tomography. Structure 19:1573–1581
103. Raines KS, Salha S, Sandberg RL, Jiang H, Rodriguez JA, Fahimian BP, Kapteyn HC, Du J, Miao J (2010) Three-dimensional structure determination from a single view. Nature 463:214–217
104. Reich ES (2011) Three-dimensional technique on trial. Nature 480:303
105. Hoppe W, Langer R, Frank J, Feltynowski A (1969) Image differentiation procedures in electron microscopy. Naturwissenschaften 56:267–272
106. Maiden AM, Humphry MJ, Zhang F, Rodenburg JM (2011) Superresolution imaging *via* ptychography. J Opt Soc Am A Opt Image Sci Vis 28:604–612
107. Dierolf M, Menzel A, Thibault P, Schneider P, Kewish CM, Wepf R, Bunk O, Pfeiffer F (2010) Ptychographic X-ray computed tomography at the nanoscale. Nature 467:436–439
108. Rossmann MG (2012) Crystallography, evolution, and the structure of viruses. J Biol Chem 287:9552–9559
109. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, Georgiev I, Kwon YD, Lee D, Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480:336–343
110. Bahar MW, Graham SC, Stuart DI, Grimes JM (2011) Insights into the evolution of a complex virus from the crystal structure of vaccinia virus d13. Structure 19:1011–1020
111. Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortin J, Martín-Benito J (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog 5: e1000491
112. Abrescia NG, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral universe. Annu Rev Biochem 81:795–822

113. Kaufmann B, Plevka P, Kuhn RJ, Rossmann MG (2010) Crystallization and preliminary X-ray diffraction analysis of West Nile virus. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:558–562
114. Kaufmann B, Nybakken GE, Chipman PR, Zhang W, Diamond MS, Fremont DH, Kuhn RJ, Rossmann MG (2006) West Nile virus in complex with the Fab fragment of a neutralizing monoclonal antibody. Proc Natl Acad Sci U S A 103:12400–12404
115. Yu X, Qiao M, Atanasov I, Hu Z, Kato T, Liang TJ, Zhou ZH (2007) Cryo-electron microscopy and three-dimensional reconstructions of hepatitis C virus particles. Virology 367:126–134
116. Badia-Martinez D, Peralta B, Andrés G, Guerra M, Gil-Carton D, Abrescia NG (2012) Three-dimensional visualization of forming Hepatitis C virus-like particles by electron-tomography. Virology 430:120–126
117. Gastaminza P, Dryden KA, Boyd B, Wood MR, Law M, Yeager M, Chisari FV (2010) Ultrastructural and biophysical characterization of hepatitis C virus particles produced in cell culture. J Virol 84:10999–11009
118. Owen RL, Axford D, Nettleship JE, Owens RJ, Robinson JI, Morgan AW, Doré AS, Lebon G, Tate CG, Fry EE, Ren J, Stuart DI, Evans G (2012) Outrunning free radicals in room-temperature macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 68:810–818
119. Murata K, Liu X, Danev R, Jakana J, Schmid MF, King J, Nagayama K, Chiu W (2010) Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. Structure 18:903–912

Further Reading

Abeyrathne PD, Chami M, Pantelic RS, Goldie KN, Stahlberg H (2010) Preparation of 2D crystals of membrane proteins for high-resolution electron crystallography data collection. Methods Enzymol 481:25–43

Abrescia NG, Grimes JM, Fry EE, Ravantti JJ, Bamford DH, Stuart DI (2010) What does it take to make a virus: the concept of the viral "self". In: Twarock R, Stockley PG (eds) Emerging topics in physical virology. Imperial College Press, London, pp 35–58

Derosier D (2010) 3D reconstruction from electron micrographs a personal account of its development. Methods Enzymol 481:1–24

Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, Jiang W, Ludtke SJ, Medalia O, Penczek PA, Rosenthal PB, Rossmann MG, Schmid MF, Schröder GF, Steven AC, Stokes DL, Westbrook JD, Wriggers W, Yang H, Young J, Berman HM, Chiu W, Kleywegt GJ, Lawson CL (2012) Outcome of the first electron microscopy validation task force meeting. Structure 20:205–214

Johnson JE (2008) Multi-disciplinary studies of viruses: the role of structure in shaping the questions and answers. J Struct Biol 163:246–253

Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. Resolution, obtained by X-ray analysis. Nature 185:416–422

Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40:191–285

Read RJ, Adams PD, Arendall WB III, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lütteke T, Otwinowski Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ, Vriend G, Zwart PH (2011) A new generation of crystallographic validation tools for the protein data bank. Structure 19:1395–1412

Steven AC, Baumeister W (2008) The future is hybrid. J Struct Biol 163:186–195
Vellieux FM, Read RJ (1997) Noncrystallographic symmetry averaging in phase refinement and extension. Methods Enzymol 277:18–53

Also especially recommended for further reading are references [7, 9, 13, 25, 27, 99, 112] *listed above*.

Chapter 8
Atomic Force Microscopy of Viruses

Pedro J. de Pablo

Abstract Atomic force microscopy (AFM) is a helpful tool to acquire nanometric-resolution images, and also to perform a certain physical characterization of specimens, including their stiffness and mechanical resilience. Besides of the wide range of applications, from materials science to biology, this technique works in a variety of conditions as long as the sample is supported on a solid surface, in air, ultra high vacuum or, most importantly for virus research, in liquids. The adaptability of this technique is also fostered by the variety of sizes of the specimens that it can dealt with, such as atoms, molecules, molecular complexes including viruses and cells, and the possibility to observe dynamic processes in real time. Indeed, AFM facilitates single molecule experiments enabling not only to *see* but also to *touch* the material under study (*i.e.*, to undertake mechanical manipulations), and constitutes a fundamental source of information for material characterization. In particular, the study of the mechanical properties at the nano-scale of viruses and other biomolecular aggregates, is providing an important set of data which help to elaborate mechano-chemical structure/function models of molecular biomachines, expanding and complementing the information obtained by other structural techniques.

Keywords Atomic force microscopy • Force • Beam deflection • Tip • Cantilever • Stylus • Topography • Mechanics • Liquids • Tip-sample dilation • Adsorption • Virus • Capsid • Virion • Deformation • Nanoindentation • Force curve • Disruption • Breakage

P.J. de Pablo (✉)
Department of Physics of the Condensed Matter, C03, Facultad de Ciencias, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: p.j.depablo@uam.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_8,

Abbreviations

AFM	Atomic force microscopy
CCD	Charge-coupled device
CM	Contact mode
DM	Dynamic mode
EDL	Electrostatic double layer
EM	Electron microscopy
HOPG	Highly oriented pyrolytic graphite
JM	Jumping mode
MVM	Minute virus of mice
STM	Scanning tunneling microscopy
UHV	Ultra-high vacuum

8.1 Introduction

The word *microscope* suggests an optical device that uses light to achieve a magnified image of a small object. Consequently, the question that arises when seeing an atomic force microscope (AFM) for the first time is: where do I have to look to see the object? In a conventional microscope (Fig. 8.1), a source emits particles, such as electrons or photons, that interact with the specimen. A detector registers and communicates this interaction to an analyzer, which processes the information received to make it comprehensible. Optical and electron microscopes fit into this source-specimen-detector-analyzer scheme.

In the optical microscope, the photons emitted by the incandescent lamp are conveniently manipulated by a system of lenses, both before and after the interaction with the specimen, finally arriving to the eyepiece where the detector (*i.e.* the eyes, film or charge-coupled device, CCD) collects the information. A typical optical microscope using visible light can reach a resolution of $\lambda/2 \sim 200$ nm (λ being the wavelength of the light).

In the case of electron microscopy (EM) (see Chap. 3), a filament emits electrons and a system of electromagnetic lenses are used to manipulate and focus the electron beam both before and after the interaction with the specimen. Afterwards the electrons are collected by a phosphorescent screen, film or CCD. In 1933 the German engineering Ernst Ruska constructed an electron microscope that exceeded the maximum resolution of an optical microscope, reaching about 1 nm.

AFM is not based in registering interactions between probe particles, such as photons and electrons, and the sample. In AFM the probe is a nanometric tip at the end of a microcantilever which mechanically palpates the object under study like a blind person uses a walking stick. In this way AFM enables the construction of a topographic image based upon recording the deflection angle of the microcantilever at each point of the studied sample. In this chapter, I describe the conceptual and

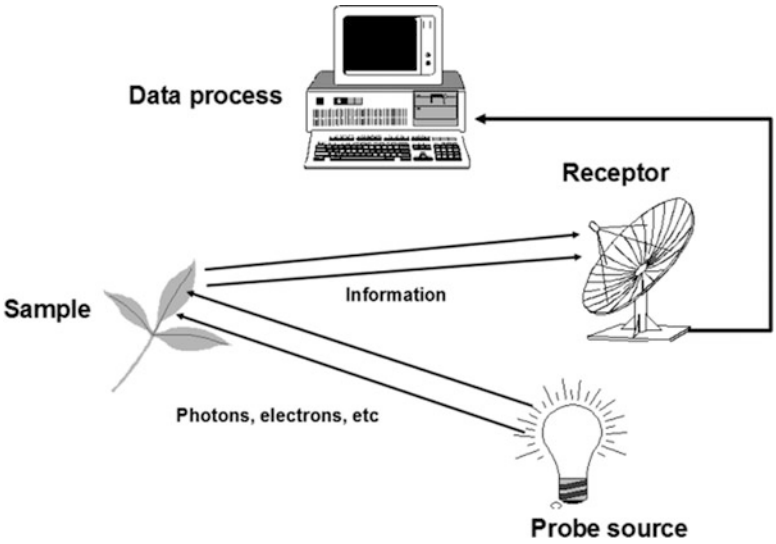


Fig. 8.1 Concept of microscopy. Any microscope requires probe source, sample, receptor and data analysis

technical bases of AFM. In particular, I refer to working methods in liquids and its applications for imaging virus particles and viral components (Fig. 8.2) In a related Chapter (Chap. 18), the application of AFM to determine the mechanical properties of virus particles is described.

8.2 Basic Concepts

In scanning probe microscopy a sharp stylus of a few nanometers in diameter, which can be considered as a probe, approaches the surface of the sample. Binnig and Rohrer [7] invented the Scanning Tunneling Microscope (STM) and received the Nobel Prize for Physics along with Ruska in 1986. The STM is based on a quantum effect (tunneling) that occurs when a sharp metallic tip is brought to a distance (z) of less than 1 nm from a conductive surface. This effect involves the flow of an electronic current (I) between the surface and the tip according to the formula $I \propto \exp(-\sqrt{\phi}z)$, where ϕ is the work-function of the metallic surface [8]. The strong dependence on the tip-surface distance can be used to obtain topographic and electronic maps of the sample by moving, (*i.e.* scanning) the tip on the surface while keeping the tip-sample distance constant through a feedback algorithm. Although this tool provides true atomic resolution in ultra-high vacuum (UHV) conditions, a mandatory requisite is that both the tip and sample should be conductive. Therefore, STM is not suitable for biological samples since these are mainly insulators, which need to be covered with a metallic layer [9].

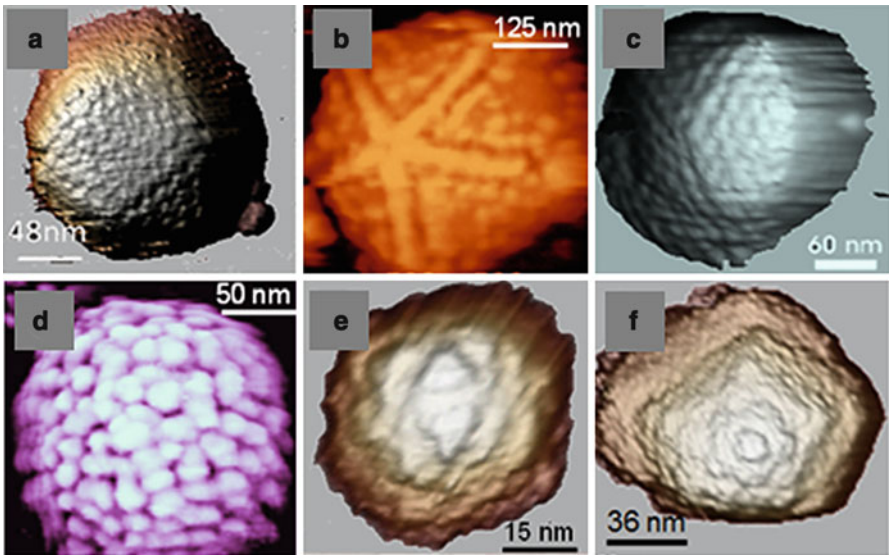


Fig. 8.2 A gallery of virus images obtained by AFM. (**a**) Human adenovirus, taken from Ref. [1]. (**b**) Giant mimivirus, taken from Ref. [2]. (**c**) Herpes simplex virus, taken from Ref. [3]. (**d**) Moloney murine leukemia virus, after taken from [4]. (**e**) MVM, taken from [5]. (**f**) bacteriophage T7, taken from [6]. All images are reproduced with permission

In 1986 Binnig, Quate and Gerber invented the AFM [10] combining the principles of both the STM and the so-called stylus profilometer [11]. In an AFM (Fig. 8.3) a sharp tip (approximately tens of nm in radius) attached to the end of a microcantilever is approached to the surface of an object by means of a piezoelectric device. As a consequence a force appears between the tip and surface that can be attractive or repulsive (see below) causing the cantilever to bend. A feedback algorithm acting on the piezoelectric device controls the relative tip-surface distance, and a topographic map of the object is obtained by scanning the surface in a plane perpendicular to the tip. The original article by Binnig et al. [10] presents the topographic profiles obtained for an insulator (ceramic). The fact that both tip and sample may be insulators is one of the main advantages of AFM, opening the door to a whole range of possibilities such as the study of biological samples including proteins, nucleic acids, membranes, cells, and biomolecular complexes such as viruses.

8.3 AFM Implementation

Although there are a variety of routes to control the deflection of the cantilever in AFM, here we will focus on the beam deflection method [12] since it is commonly employed when working with biological samples. The beam deflection system

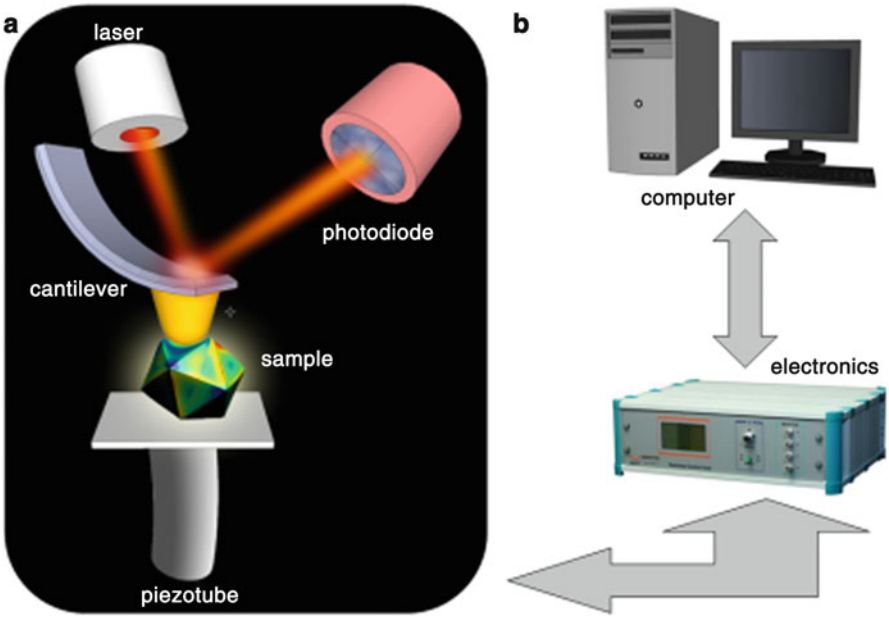


Fig. 8.3 Principle of atomic force microscopy. The software running on the computer controls the electronics (**b**), which gives and receives information to and from the AFM head (**a**)

involves focusing a laser beam on the end of the cantilever and collecting the reflected light with a photodiode (Fig. 8.3a). As a consequence, any bending of the cantilever will affect the position of the reflected laser spot on the photodiode. A normal bending will originate a so-called normal force signal, $F_{n,}$ on the photodiode sectors, whereas a lateral torsion will result in the so-called lateral force, F_l.

The core of an AFM is the head (Fig. 8.3a), where the beam deflection system is integrated along with the piezoelectric device (piezo-tube) that moves the sample in all three directions (x, y, z). In Fig. 8.3a an AFM head configuration is shown where the tip is fixed and the sample is moved by the piezo-tube. Another configuration known as "stand alone", fixes the sample and the tip makes the scanning movement controlled by a piezo-tube that carries the cantilever. The electronic components receive the signals coming from the photodiode, mainly F_n and F_l, and provide high voltages (up to hundreds of volts) to move the piezo-tube to which the sample (or cantilever) is attached. The computer is in charge for managing the data and calculating all the parameters required to move the piezo-tube in a convenient way by using the electronics (Fig. 8.3b). Details on the different components of a typical AFM and AFM operation are given in the following subsections.

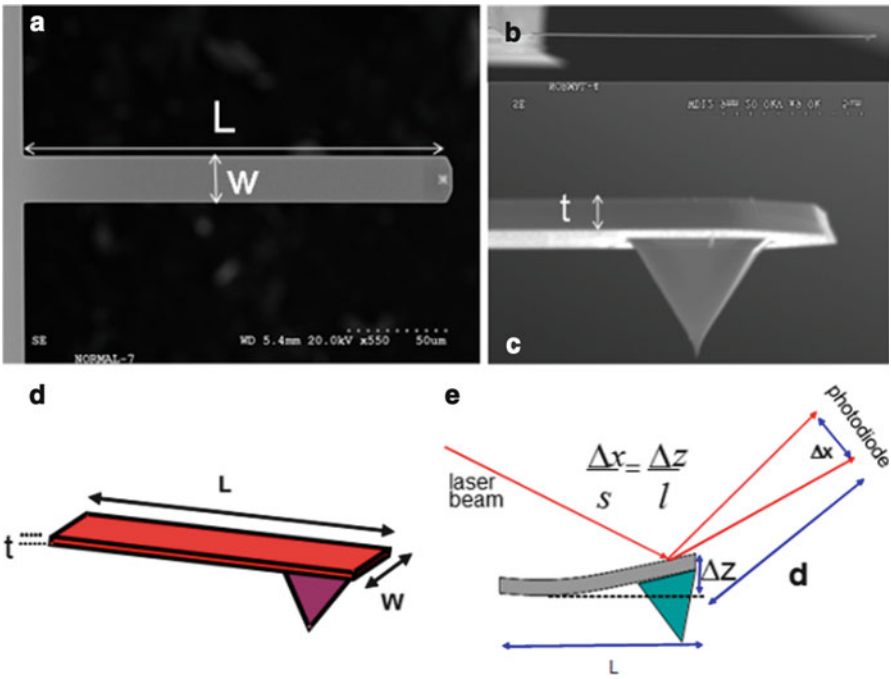


Fig. 8.4 AFM cantilevers. (**a**) and (**b**) respectively show top and side views of a rectangular cantilever of 150 μm in length (L) and 25 μm in width (w) attached to a chip. (**c**) presents a close-up image of a pyramidal tip, and a cantilever with a thickness of 150 nm. (**d**) sketches the geometry of a rectangular cantilever. The cartoon in (**e**) indicates the vertical resolution of the bending cantilever as a function of several geometric parameters

8.3.1 Cantilevers

Integrated tip and cantilever assemblies can be fabricated from silicon or silicon nitride using photolithographic techniques. More than 1,000 tip and cantilever assemblies can be produced on a single silicon wafer. The cantilevers can be rectangular (Fig. 8.4a) or V-shaped, and they typically range from 60 to 200 μm in length, 10–40 μm in width, and 0.3–2 μm in thickness. The usual tip radius is about 20 nm although lower diameters can be obtained (Fig. 8.4b). The mechanical stiffness of the cantilever (see Chap. 18) is given by its spring constant k, which normally ranges between 0.03 N/m and 40 N/m, and strongly depends on the cantilever's dimensions. For example, for a rectangular cantilever $k = \frac{EW}{4}\left(\frac{t}{L}\right)^3$, where E is the Young Modulus (a measurement of the intrinsic elasticity of a solid material, see Chap. 18) of the cantilever, and W, t and L are the cantilever width, thickness and length respectively (Fig. 8.4d). While W and L can be fairly precisely known, t is always difficult to measure. As a consequence, the manufacturers normally provide the cantilever spring constant with an error of 10–30 % and therefore the user should calibrate each cantilever [13]. The spring constant k is

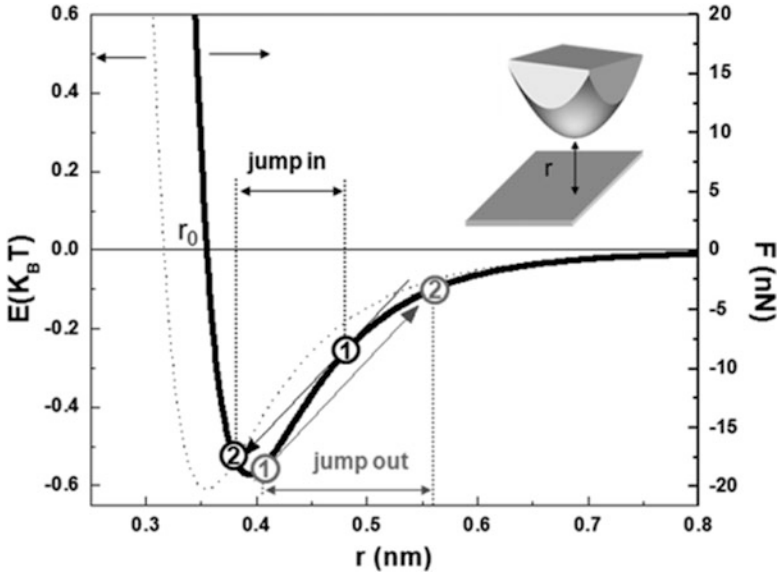


Fig. 8.5 Tip-surface interaction. *The dotted line* represents a Lennard-Jones potential as a function of the distance between atoms, r, mimicking the tip-surface interaction (represented in the *inset*). The *solid line* is the interaction force obtained from the potential. The numbers *inside the circles* indicate the cantilever instabilities (see text for a complete description)

used to calculate the force applied on the cantilever as a function of bending Δz using Hooke's law, *i.e.*, $F = k \times \Delta z$ (Fig. 8.4e).

The vertical resolution of the cantilever, Δz, strongly depends on the noise of the photodiode, since it defines the minimum significant displacement Δx, on the detector (Fig. 8.4e). A typical cantilever of 100 μm length has about 0.1 Å resolution assuming a signal-to-noise ratio around 1 [13].

Another important parameter that can have influence on the vertical resolution is the thermal noise [14]: the cantilever oscillates at the resonance frequency with an amplitude $\Delta z = \sqrt{\frac{k_B T}{k}}$, where k_B is the Boltzman constant, T is the absolute temperature and k the cantilever spring constant. For example, at room temperature, there is a noise of about 5 Å for a cantilever of spring constant 0.02 N/m.

8.3.2 Interaction Between the AFM-Tip and the Sample Surface

In order to understand the interaction between tip and sample we shall refer to potentials rather than forces. Physicists prefer this approach as potentials are scalar and therefore easier to deal with compared to vectors such as forces. For this purpose let us graphically depict the Lennard-Jones potential in Jules (J) between two atoms separated by r meters (m) in vacuum [15] (dotted curve in Fig. 8.5).

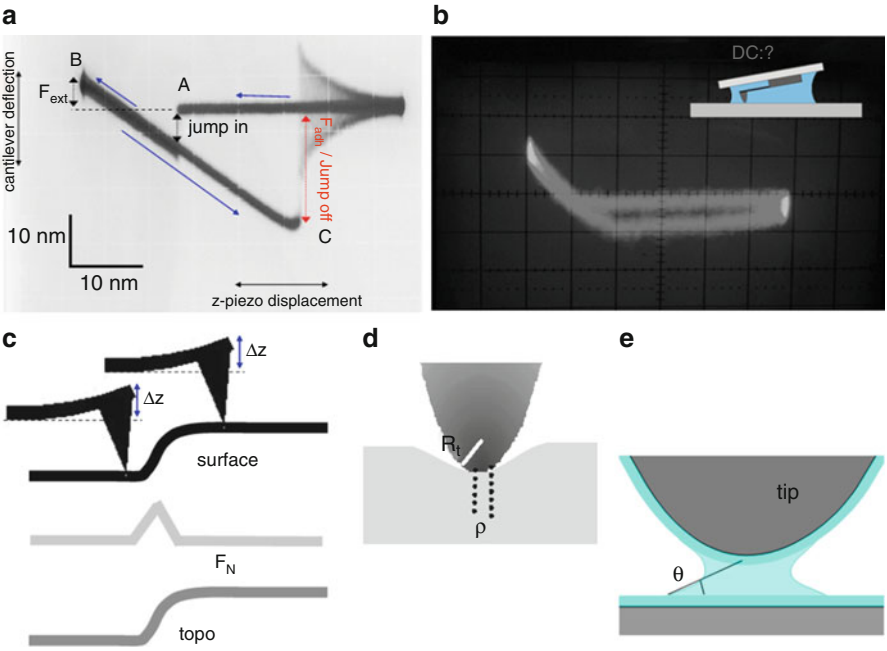


Fig. 8.6 Force vs. distance experiments, feedback and lateral resolution. (**a**) and (**b**) show oscilloscope snapshots of force vs. *z*-piezo displacement curves in air and water, respectively. (**c**) presents the contact mode showing the variation of F_n (*light gray*), and *z*-piezo tube voltage (topography, *dark gray*) as a function of a step (*black*). In (**d**) the geometric features of the tip-surface contact are depicted. (**e**) shows a cartoon illustrating the capillary forces between tip and surface (see text for a complete description)

$U(r) = -A/r^6 + B/r^{12}$, where r is the tip-sample distance, $A = 10^{-77}\ J\ m^6$ and $B = 10^{-134}\ J\ m^{12}$.

A rough approximation to the AFM tip-sample interaction is to consider the approach of two such atoms where, as depicted in the inset in Fig. 8.5: the lower part shows a solid surface and the upper part shows the apex of a sharp tip is moving downwards. An interaction force $F(r) = -\frac{dU}{dr}$ will occur and it can be seen that this force is attractive ($F < 0$) when $r > r_0$ or repulsive ($F > 0$) when $r < r_0$, (solid line curve in Fig. 8.5). These regions define the attractive and the repulsive regimes of operation, respectively.

Now let us consider the role of these regimes in a force vs. distance (FZ) AFM experiment that involves the tip approaching to the surface (Figs. 8.5 and 8.6). The experiment starts with the tip situated far from the surface in the attractive regime. As the tip is approaching and as soon as the gradient (*i.e.*, the slope) of the force equals the cantilever spring constant, the tip jumps to the surface from black point 1 to black point 2 in Fig. 8.5 (both connected by the slope). This effect is seen in the FZ of Fig. 8.6a at point A like a sudden jump of the cantilever deflection (vertical scale). In this way the tip establishes mechanical contact with the surface and it rapidly enters the repulsive regime ($F_N > 0$). The *z*-piezo (horizontal scale) is

elongated until a given F_N or deflection value is reached and then stops (point B in Fig. 8.6a). The external loading force F_{ext} can be calculated as the difference between the zero deflection position (*i.e.* before the jump to contact in point A in Fig. 8.6a) and the deflection at point B in Fig. 8.6a. Since the vertical scale is about 10 nm per division and the cantilever spring constant $k = 0.1\ N/m$ we have $F_{ext} \sim 0.5div \times 10nm/div \times 0.1nN/nm = 0.5nN$.

Subsequently, the *z*-piezo retrace cycle starts, and the tip is released from the surface at point C in Fig. 8.6a (*i.e.*, once more where the derivative of the tip-surface force equals the cantilever spring constant) jumping from the grey point 1 to grey point 2 in Fig. 8.5, following the grey arrow. The cantilever deflection jumps-off to zero with a dampened oscillation. This jump-off is known as the adhesion force F_{adh} (in Fig. 8.6a $F_{adh} \sim 2div \times 10\ nm/div \times 0.1nN/nm = 2\ nN$). The total force at point B in Fig. 8.6a is the sum of F_{ext} and F_{adh}, (*i.e.*, 2.5 *nN*). It is interesting to stress that no matter how small F_{ext} is, the total force applied to the surface will always be at least F_{adh}.

8.3.3 Contact Mode (CM)

CM is the simplest operational method used for AFM and it was the first to be developed [10]. Here the tip is brought into contact with the surface until a given deflection in the cantilever (F_n) is reached; the tip then scans a square area of the surface to obtain a topographic map. By elongating or retracting the *z*-piezo, the feedback algorithm tries to maintain the cantilever deflection constant by comparing the F_n signal with a set point reference value established by the user. Topographic data are obtained by recording the *z*-piezo voltage that the feedback algorithm is applying to correct the cantilever deflection at each position on the surface. Since the *z*-piezo is calibrated, voltages are transformed into heights and a topographic map is obtained. Let us consider a simple example where the tip is scanning a step (black line in Fig. 8.6c) with $F_n = k \times \Delta z$. When the cantilever moves to the upper part of the step, it undergoes a deflection greater than Δz. Therefore the feedback algorithm retracts the z-piezo in order to achieve the same deflection Δz as when the tip was in the lower part of the step and, as a consequence, a topographic profile of the step is obtained (blue line in Fig. 8.6c). On the other hand F_n varies at the step that is corrected by the feedback, which can be observed as a peak in the deflection signal (red line in Fig. 8.6c). The latter is known as the constant deflection mode or constant height mode, since the *z*-piezo elongation is not modified and a map of the changes in F_n is obtained. The reader is encouraged to reproduce the profiles of Fig. 8.6c when the tip goes down the step.

Let us consider now the lateral resolution that can be achieved in contact mode. We can make an estimate of this parameter by applying the Hertz theory [16], which accounts for the deformation of solids in contact. Once the tip is in contact with the sample, the radius ρ of the tip-surface contact area is given by (Fig. 8.6d):

$$\rho = \left(\frac{3 \cdot F \cdot R}{4 \cdot E^*}\right)^{1/3}$$

Where F is the applied force *i.e.* F_n; E^* is the effective Young modulus expressed by

$$\frac{1}{E^*} = \frac{1-\nu_t^2}{E_t} + \frac{1-\nu_s^2}{E_s}$$

With E_t, ν_t and E_s, ν_s being the Young Modulus and Poisson ratio for the tip and sample respectively. R is the effective radius expressed as a combination of the tip R_t and sample radius R_s,

$$\frac{1}{R} = \frac{1}{R_{tip}} + \frac{1}{R_{sample}}$$

It is worth to stress that the main condition for Hertz model applicability is that the radius of contact ρ has to be much less than the effective radius R: ρ<<R. This approximation roughly implies that the deformation is smaller than the dimensions of the tip-sample system [16]. It is important to recall that this assumption does not limit Hertz model to deformations of flat samples. In the case of metals $E = 100\ GPa$, $\nu \sim 0.5$ and the mechanical thermal noise of the cantilever is 10 pN, with $R_t \sim 20\ nm$, the radius of contact ρ is about 0.15 nm, which implies atomic resolution. However, the adhesion force in air (see below) is about 5 nN, increasing ρ to about 1 nm. Atomic resolution can be achieved by either working in liquids [17] or in UHV conditions [18], where even individual atoms can be chemically identified [19].

CM is not very frequently used to image isolated biological materials such as virus particles. The very high lateral forces that are applied to the surface [20], which can damage the sample, present an important problem when using this mode. This is especially important for single mechanically delicate biomolecules adsorbed onto a surface.

8.3.4 Dynamic Modes (DM)

In DM operation [21] the cantilever is oscillated near to or at its resonance frequency ($\omega_0 \propto \sqrt{E}t/L^2$ for a rectangular cantilever). As the tip approaches the sample the oscillating amplitude decreases until it establishes contact with the sample (Fig. 8.7a), following a similar cycle to that in Fig. 8.6a but now with oscillation. Therefore the feedback loop involves the amplitude rather than the F_n and, by keeping the oscillation amplitude constant, a topographical map of

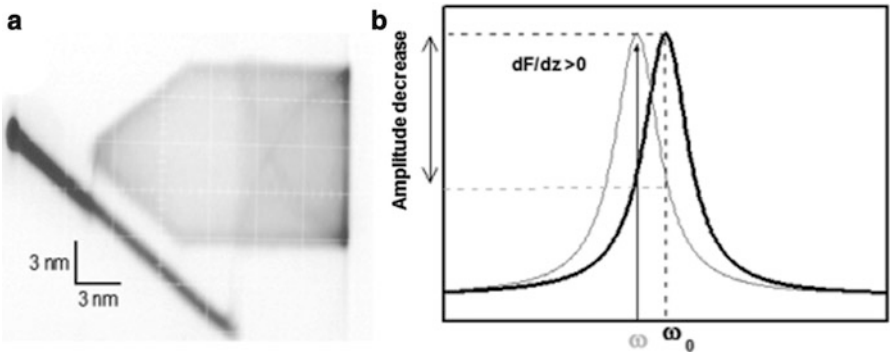


Fig. 8.7 Dynamic modes. (**a**) depicts the force vs. Piezo-tube z displacement curve of an oscillating cantilever. (**b**) illustrates the reduction of the resonance frequency of the cantilever from ω_0 to ω as the tip approaches to the surface (see text)

the object can be obtained. The amplitude is reduced because the resonance frequency (ω) changes with the tip-sample distance (z) and with the tip-sample interaction force F_{ts} according to $\frac{\Delta\omega}{\omega_0} \propto \left(\frac{1}{2k}\right)\frac{dF_{ts}}{dr}$ [22], thereby decreasing with attractive forces. The new resonance frequency ω, is lower than ω_0 and, since the cantilever is still oscillating at ω_0, the cantilever amplitude decreases (Fig. 8.7b).

While CM can damage isolated biological samples on surfaces, DM operated in non-contact mode [22] does not apply high dragging forces and is commonly used to image weakly attached molecules, including delicate biomolecules, in air. In addition to topographic maps, DM can be used to obtain other information on the sample, including a phase map (the time difference between the excitation and the response of the cantilever measured for each point in the sample), which in air carries information on the different materials composing of the sample [22].

When DM is used in liquid the situation changes completely, as the viscosity of water reduces the resonance frequency about fourfold and the quality factor (a measure of the cantilever damping) of the oscillation is reduced from ~100 to 10. When oscillating the cantilever in liquid a mechanical contact between tip and sample is established, thereby applying lateral and normal forces that may damage the specimen, especially when delicate biological objects are imaged [23]. Thus, while DM can be advantageously used to image virus particles in air, its use to image them in more physiological (liquid) conditions may be limited, unless fixation agents such as glutaraldehyde are used [24].

8.3.5 Jumping Mode (JM) or Pulse Force Mode

JM, also called pulse force mode [25, 26] is a variation of CM where lateral tip displacement occurs when the tip and sample are not in mechanical contact, thereby avoiding to a large extent shear forces and the corresponding damage to the

tip-sample system. JM performs a FZ curve (Fig. 8.6a) at every point of the sample (pixel in the image), moving the tip to the next point at the end of each cycle when the tip and sample loose contact. Feedback is engaged at point B in Fig. 8.6a, moving the piezo-tube along the z-axis in a convenient way to maintain a constant deflection or loading force F_{ext}.

JM routinely produces adhesion force maps, which provides compositional or geometrical information of the surface [27]. The adhesion force between the tip and the surface can be described as $F_{adh} = 4\pi R\gamma_L \cos\theta + 1.5\pi\Delta\gamma R$, where γ_L is the water surface tension and θ the angle of the water meniscus present between the tip and surface; R is the effective radius (described above) and $\Delta\gamma$ the tip-surface energy difference. Since the first term accounts for the water meniscus between tip and surface (Fig. 8.6e), it mainly informs about the hydrophobicity of the sample. A rough estimation of adhesion force in air at room temperature results in ~7 nN for a tip radius of about 20 nm. The second term depends mainly on the tip-surface geometry. We denote the importance of the first term by comparing the FZ curves taken from glass in both air (Fig. 8.6a) and water (Fig. 8.6b): in liquid the adhesion force is almost absent, since there is not water meniscus between the tip and sample, although some hysteresis appears in the FZ due to the dragging of the water on the cantilever. In our view, JM is a mode of choice for analyzing virus particles in liquid in close to physiological conditions, because it minimizes lateral forces and enables a precise control of the force applied [28].

8.3.6 Tip-Sample Geometrical Dilation

All of the operational modes described above give rise to an effect which is inherent to the very nature of the AFM approach: the so-called tip-sample geometrical dilation. Because of this geometry-based effect, the topography of an object in an AFM image appears as laterally expanded (see Fig. 8.8). AFM users must understand and control this artefact as much as possible in order to preclude any misunderstanding of the data obtained.

When surface asperities are of dimensions comparable to the tip radius, which is very common with many samples imaged by AFM, including virus particles, the size of the tip plays an important role on image definition and the apparent image lateral expansion due to the dilation of certain features of the image by the finite tip size [29]. Figure 8.8a, b present AFM images of single-walled carbon nanotubes on silicon oxide; these images were obtained using tips of different radii, lower in Fig. 8.8a and higher in Fig. 8.8b. A carbon nanotube is rolled-up graphene in the form of a cylinder of a few nanometers in diameter. Fig. 8.8c shows the geometric dilation effects with the different tip radii: a single topography scan line across a carbon nanotube results wider for Fig. 8.8b (black) than for Fig. 8.8a (red). This effect can be seen to arise from the geometrical considerations shown in Fig. 8.8d, and the tip radius r_t can be calculated as $r_t = b^2/2h$, where b and h are

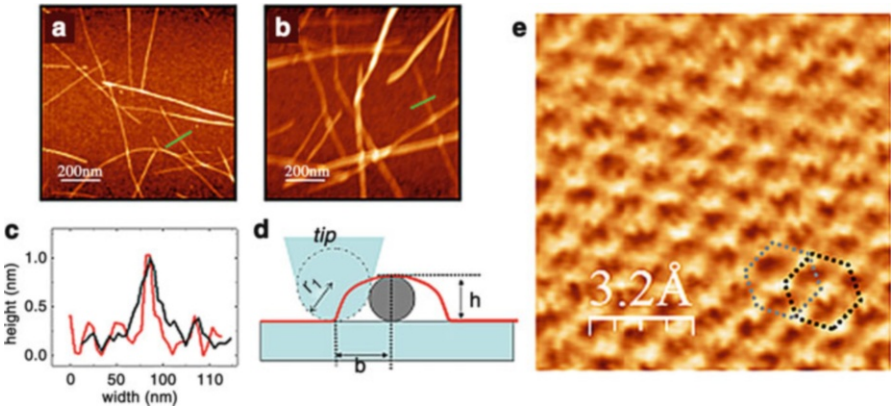


Fig. 8.8 Tip-sample dilation and atomic corrugation in HOPG. (**a**) and (**b**) show AFM images of the same samples of carbon nanotubes on silicon oxide. The topographic profiles obtained along the *green lines* in (**a**) and (**b**) show different widths due to dilation in (**c**). (**d**) illustrates the geometric parameters in the dilation process. The *red line* depicts the dilated section of the carbon nanotube section (*grey circle*). (**e**) presents the atomic corrugation of HOPG, where the two carbon layers results in two hexagonal networks

the width and the height of the sample, respectively. By using this formula, it can be calculated that the topographies of Fig. 8.8a, b have been obtained using tips with a radius of 15 nm or 70 nm, respectively. We can conclude that the sharper the tip, the better the image (which will become less laterally expanded and more defined). A popular and impressive experiment frequently used to teach AFM is to image highly oriented pyrolitic graphite (HOPG) in air [30], where the elastic deformation of the tip-sample contact plus the dilation effects result in atomic corrugation (Fig. 8.8e), although atomic defects are not visualized [31]. During imaging by AFM of virus particles and other biomolecular complexes, the height of the object at each defined point can be determined with subnanometer precision; however, the tip-sample dilation effects are important when considering lateral image dimensions and the object fine details.

8.4 Imaging Viruses and Other Biological Objects

8.4.1 Nucleic Acids, Proteins and Biomolecular Complexes

During the virus life cycle, many interactions between viral components and between viral and cellular components occur. We will review here some applications of AFM for imaging biological molecules, biomolecular complexes and their interactions.

AFM imaging in biology can be carried out in air or in liquid. Some relevant examples of AFM imaging of biomolecules in air are described first. For example, DM is frequently used for imaging DNA on a mica surface. As both DNA and mica are negatively charged, $MgCl_2$ is added to the DNA solution; the Mg^{2+} ions are sandwiched between DNA and mica, resulting in the adsorption of DNA molecules on the surface. Subsequently, the sample is dried out and DM is used to image the DNA molecules on the surface. Some structural and physical aspects of the DNA molecule have been the focus of much research using AFM [32]; in addition, an important biological application of AFM is to investigate the binding of proteins to DNA [33–35]. In this kind of single-molecule experiments the researcher is not probing the average result of the bulk reaction, but rather the action of single protein molecules on a single DNA molecule. Once the protein solution is pipetted into the DNA solution, the reaction starts and the DNA-protein complex is adsorbed onto the mica at the appropriate time and air-dried. Therefore, on the surface there is a snapshot of the process taking place between proteins and DNA, and the topography provides single molecule information of the protein-DNA complex.

However, in biology we are particularly interested in imaging in a liquid milieu in close to physiological conditions. This is an important point since biomolecules are normally fully functional only in aqueous solutions. In such a case, several particular aspects should be considered. First, a strong enough attachment of the biomolecule to the supporting surface is required to achieve good resolution during the AFM imaging process in liquid. Although biomolecules can be covalently linked to a chemically modified surface [36], covalent modifications could potentially damage the biomolecules, and therefore this treatment tends to be avoided. Fortunately, physisorption is usually sufficient and thus, the specimens can be directly adsorbed to a solid surface in a physiological buffer. The relevant forces that drive the physisorption process are the van der Waals force, the electrostatic double-layer force (EDL force) and the hydrophobic effect [37]. Unlike the van der Waals interaction, the EDL force depends strongly on the concentration and valence of charged solutes, as well as the surface charge density of both surface and specimen. The EDL force between two equally charged surfaces is repulsive and hence opposite to the van der Waals attraction [38]. One of the most extended AFM applications is to image 2-dimensional protein crystals in liquid, as well as lipid membranes, by using contact mode. Biological macromolecules become attached to the surface (mica, silicon, gold, glass, etc.) when there is a net attractive force between them and the surface pulling their surfaces into contact. This force per unit of surface (pressure) can be estimated to the light of the DLVO theory [15] like the sum of the electrostatic force between surface and molecule F_{el}, and the van der Waals interaction F_{vdW}.

$$F_{DLVO} = F_{el}(z) + F_{vdW}(z) = \frac{2\sigma_{surf}\sigma_{sample}}{\varepsilon_e \varepsilon_0} e^{-z/\lambda_D} - \frac{H_a}{6\pi z^3}$$

where z is the distance between the surface and specimen; σ_{surf} and σ_{sample} are the charge densities of surface and specimen, respectively; ε_e is the dielectric constant

of the electrolyte; ε_0 is the permittivity of vacuum; λ_D is the Debye length, that depends on the electrolyte valence [37]; and H_a the Hamaker constant. The adsorption of a sample onto freshly cleaved mica (atomically flat) can be manipulated by adjusting both the ion content and the pH of the buffer solution.

AFM in liquid is being increasingly used in molecular and cell biology [39–44]. A major application is to produce topographic images of single biomolecules and, especially, larger biomolecular complexes, lipid membranes and their embedded proteins, and whole cells in close to physiological conditions. Furthermore, by introducing variations in the conditions (*i.e.*, pH changes, addition of other biological molecules, etc.) the resulting changes in the general surface structure of the biomolecule or biological object of interest can be identified. Interestingly, in this kind of set-up the cantilever can be used also to apply forces that trigger conformational changes in single proteins molecules such as in the chaperonine GroEL; the changes thus induced may mimic those that occur under the action of physiological agents and can be revealed in real-time by AFM imaging [45].

AFM is being used also to visualize single proteins doing mechanochemical work. DM in liquid is generally used for this purpose since it is faster [46]. Maximum peak forces of a few nN are applied [23], in relation to the stiffness of the sample [47], and, although these forces could in principle be damaging to the sample, they are applied for very short periods of time such as 10 % of an oscillating period (*i.e.* for cantilever with a resonance frequency of 10 kHz in liquid, the forces are applied during 10 μs every 0.1 ms). Using this method, it has been possible to visualize, for example, the activity of RNA polymerase on DNA [48] or the conformational changes in a DNA-repair complex upon binding DNA [49]. Further high speed AFM developments have enabled to monitor the activity of single molecular machines in real-time [50].

In addition to imaging, AFM is also being applied in biology to measure the forces that participate in interactions between and within single biomolecules. AFM and optical tweezers (see Chap. 9) are major single-molecule techniques contributing to a better physics-based understanding of protein folding and unfolding, protein-ligand recognition, macromolecular assembly and disassembly, cell adhesion, etc. (*e.g.*, see references [32–35, 39–44, 46].

8.4.2 Virus Particles

Structural and chemico-physical characterization has been critical to understand the biology of viruses. X-ray crystallography (Chap. 4) and EM (cryo-EM) (Chap. 3) techniques have traditionally been used and provide direct three-dimensional structural information on whole virus particles, allowing both the interior and the surface of the virus to be visualized at high resolution. However, a limitation of these techniques is that they are both averaging ("bulk") techniques and thus they present an average time and space model of the entire population of particles in the crystal or of the many particles selected for cryo-EM image averaging. As such, these

techniques generally provide limited information on possible structural differences between individual particles in the population that distinguish them from the average structure. For this reason, the beautifully symmetrical and apparently perfect models of larger viruses derived from these techniques may be somewhat deceptive, as they may not be entirely representative of the individual viruses. Therefore, the imaging of individual virus particles may provide unique insights on the structural biology of viruses and should be considered as an important part of virus characterization [51]. In addition, AFM allows imaging viruses in liquid state and following in real-time the structural changes of individual virus particles in response to different conditions and added agents, provided the changes are not too fast (see also Sect. 6). This may enable the analysis, for example, of the stages of virus assembly or disassembly [52]. Virus particles can be disrupted upon indentations above the rupture strength [3, 53, 54], and mechanical fatigue below the rupture can be used to monitor in real time the intermediate stages of adenovirus disassembly [1]. Moreover, AFM can be used to determine some physical properties of virus particles including mechanical features (see Chap. 18). Finally, AFM enables to image and quantitate the interactions of single virus particles with other molecules and even host cells [55].

Since individual virus particles adsorb weakly to typical surfaces, they are prone to undesired modifications by lateral forces in CM or DM operation, as they are not held by a surrounding neighbourhood, unlike a protein in a 2-dimensional crystal. The typical approach for surpassing this limitation is the use of structural fixation agents, such as glutaraldehyde. In such conditions AFM may provide images whose resolution is comparable to that of some EM images [56], although still far less than that of the best structural models of virus particles recently obtained by cryo-EM (see Chap. 3). Nevertheless, since glutaraldheyde structurally reinforces the specimens subjected to study [57, 58], it precludes any characterization of processes or properties of intact native viruses, such as disassembly or mechanical features [59].

Chemical fixation is, however, not required in many instances to obtain good images or probe mechanical properties of viruses. The adsorption of viruses on solid surfaces is based on the same principles than those involved in the adsorption of DNA and proteins (Sect. 8.4.1). Although the typical approach implies the convenient sylanization of glass surfaces [60], treated mica [61] and Highly Oriented Pyrolitic Graphite (HOPG) (21) have also proved as adequate substrates for virus attachment.

In general, JM has been the preferred mode to analyze unfixed viruses, since the loading forces can be accurately controlled and lateral forces are minimized. Firstly, let us exemplify the tip-virus geometrical dilation with a prolate icosahedral virus. Figure 8.9a shows a topographic image of a bacteriophage ϕ29 virion. Its comparison with the EM model [62] is shown in Fig. 8.9b: AFM enables a reasonable structural delineation of the viral head, the neck region and the tail. When the tip-sample geometrical algorithm [29] implemented in the WSxM software [63] is applied, one observes (Fig. 8.9b) how the EM model (light contour) is dilated to the black contour in the AFM image (see Sect. 3.6). Figure 8.9c demonstrates how the geometrical dilation of the image of a ϕ29 virion adsorbed

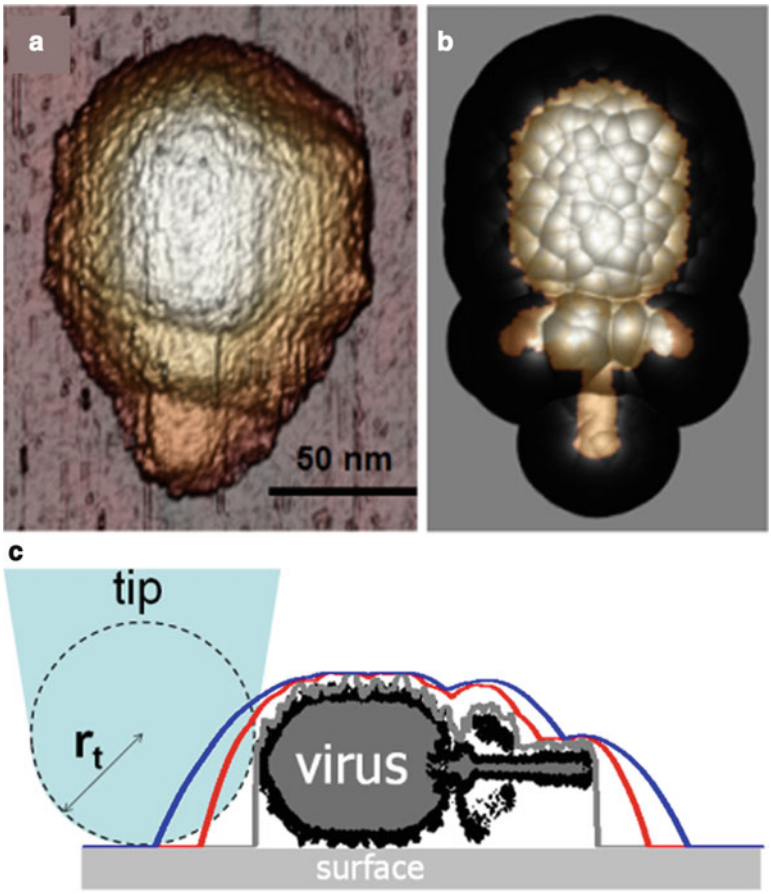


Fig. 8.9 Dilation effects in bacteriophage ϕ29 AFM images. (**a**) shows the topography of a single phage ϕ29 particle. The *bright color* image in (**b**) represents a EM structural model of ϕ29, and the *black area* indicates the dilation effects corresponding to a tip of 10 nm in diameter. The *cartoon* in (**c**) indicates the dilation effect as a function of the tip size

on a surface increases with the tip radius. A 0.5 nm-radius tip produces the grey color profile, which accurately follows the virus shape. In turn, tips of 5 and 10 nm in radius (red and blue lines, respectively) show an evident dilation and also blurring of the structural details in the virion image.

In the following, let us consider some geometrical aspects of the adsorption of icosahedral viruses (see Chap. 2) to a solid surface. For instance, the roughly spherical minute virus of mice (MVM) has a capsid of icosahedral symmetry with conspicuous protrusions at the threefold symmetry axes. Single MVM particles adsorb to the solid surface in different orientations, which frequently leaves a threefold, twofold or fivefold symmetry axis close to the top of the particle (see Figs. 8.10a–c, respectively). The possibility to identify the orientation of single

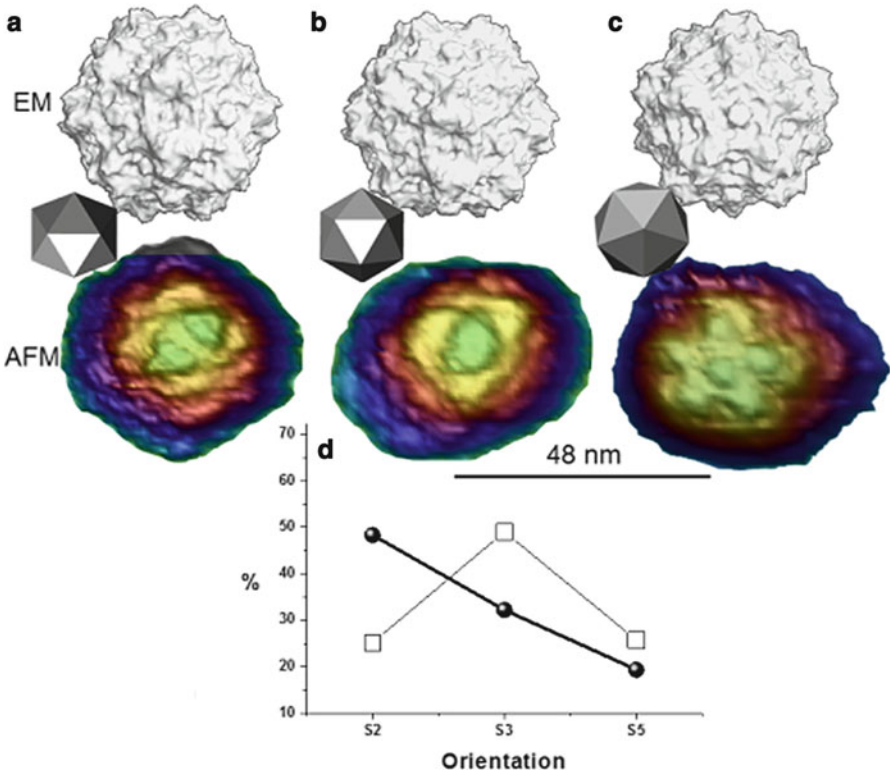


Fig. 8.10 AFM imaging of MVM. (**a**), (**b**) and (**c**) show the topographies corresponding to single MVM particles oriented with a twofold, threefold or fivefold symmetry axis on top, respectively. (**d**) represents the percentage of adsorption of each symmetry axis (*squares*) and the theoretical expectations if the adsorption through edges, faces and vertices of a perfect icosahedron were equiprobable

virus particles by resolving protrusions or other structural details allows not only a rather detailed characterization of their topography, but also the determination of local physical properties. For example, by making nanoindentations on single MVM particles adsorbed in different orientations and obtaining FZ curves, it has been shown that the viral single-stranded DNA genome is responsible for an anisotropic mechanical stiffening the virion relative to the DNA-free capsid [64], something that could be important for preserving viral infectivity. Moreover, it was possible to selectively disrupt these DNA-protein interactions or replace amino acid residues in selected regions of the MVM capsid and thereby engineer virus particles with altered mechanical properties [5] (see Chap. 18). Human adenovirus represents another example in which the orientation of single particles imaged by AFM could be determined on the adsorption of icosahedral viruses on a surface (Fig. 8.11). In this case, viruses attach to mica previously treated with $NiCl_2$ [28]. Comparison among

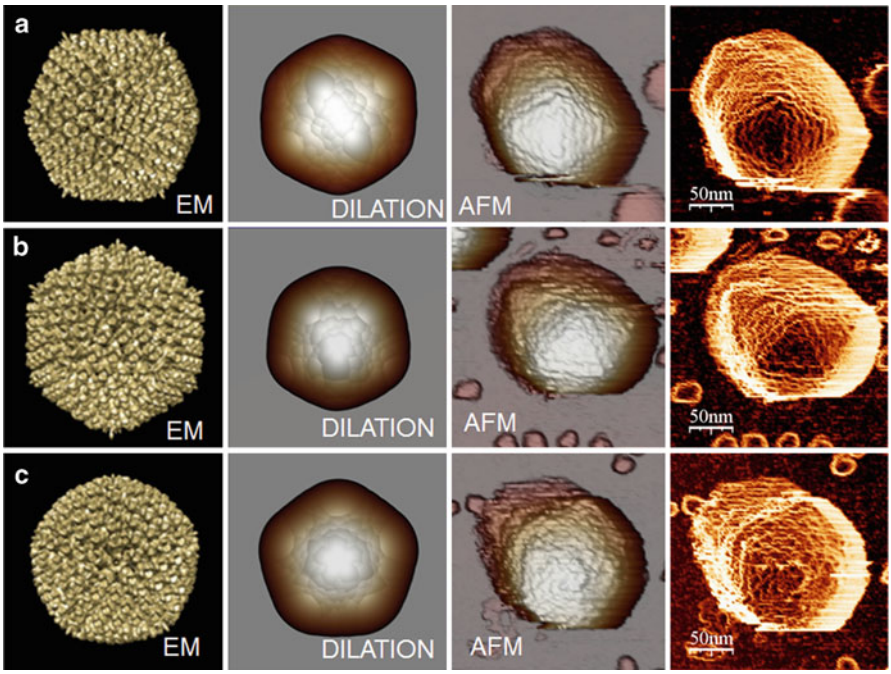


Fig. 8.11 AFM imaging of human adenovirus. (**a**), (**b**) and (**c**) show the topographies corresponding to single adenovirus particles oriented with a twofold, threefold or fivefold symmetry axis on top, respectively. AFM images are compared with EM and EM dilated structural models. The *right column* shows AFM topographic images that have been filtered to enhance the borders by obtaining the cosine of the angle between the normal vector of the surface and the normal direction of the paper sheet

EM data, EM geometrically dilated images and AFM images enabled a rapid identification of the particle orientation (see Figs. 8.11a–c for particles oriented along a twofold, threefold or fivefold symmetry axis, respectively).

It may be instructive to quantify the adsorption geometries found in MVM particles (Fig. 8.10d). Black circles represent the orientation likelihood of an icosahedron on a surface provided that the adsorption through vertices (fivefold symmetry), edges (twofold symmetry) and faces (threefold symmetry) is equally probable. If such were the case, a twofold symmetry orientation would exhibit the highest adsorption probability because in an icosahedron there are more edges (30) than faces (20) and vertices (12). On the other hand, empty squares depict the experimental results, which prime the threefold symmetry axis adsorption with almost a 50 % of the adsorbed particles. Since the threefold orientation maximizes the icosahedron area of contact, in contrast with edges or vertices, the particle-surface adsorption forces may depend strongly on the virus-surface contact area (Fig. 8.10d).

8.5 Understanding Viruses: Some Major Contributions of AFM

AFM enables new possibilities for studying virus particles, complementing classic structural approaches such as EM and X-ray crystallography. First, AFM may routinely work with individual viral particles. Second, they can be studied in liquid phase, in conditions which are closer to the *in vivo* environment. Although the resolution of AFM images is not comparable to that of structural models obtained by cryo-EM or X-ray crystallography, AFM enables the identification of virus elements which are not symmetrically ordered. Third, AFM may eventually allow the study in real-time of dynamic processes including the assembly, disassembly and conformational rearrangements of virus particles. Fourth, AFM has opened the possibility to investigate previously unexplored physical properties of virus particles, which may provide novel insights into the relationships between virus structure, properties, and functions. Along this line, AFM offers the possibility of probing the mechanical stiffness of virions and capsids and of selectively disrupting elements in individual virus particles (see Chap. 18). These experiments may contribute to understand virus assembly and disassembly, stability and dynamics.

8.6 Perspectives

Some of the current developments in AFM imaging aim to increase the data acquisition speed (a typical AFM image takes a few minutes to be acquired), in order to directly visualize the dynamics of relatively fast biological processes in real time, including conformational changes and the assembly or disassembly of virus particles. For example, AFM imaging at high video rates (80 ms/frame) has been already used to show conformational changes of single myosin V proteins on mica [65]. The methodology to achieve such a high speed AFM involves several technical developments, including the use of very soft cantilevers with a high frequency of resonance in liquids, which is achieved by decreasing the cantilever thickness and reducing the cantilever width and length proportionally.

Mechanical properties of biological objects are generally analyzed in indentation experiments [60]. However, there is also the possibility of using DM operating in liquids to extract information about mechanical properties of virus particles and other biological objects (*e.g.*, their stiffness), by mapping the phase lag [66].

The development of non-invasive AFM imaging techniques where sample destruction is minimized is also important. For example, frequency modulation AFM [67, 68] is a dynamic technique where small forces of a few tens of pN only can be applied to the surface. This promising technique is based on the use of three simultaneous feedbacks. A phase lock loop ensures that the cantilever is always in

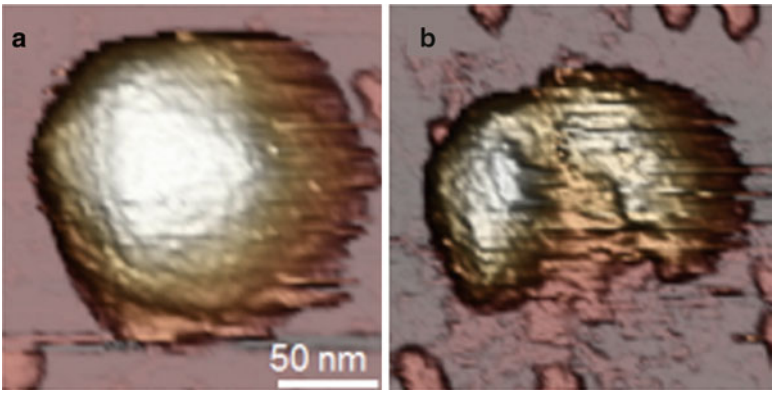


Fig. 8.12 Breaking a virus. (**a**) and (**b**) show a single human adenovirus particle before and after the breakage provoked by pushing the virus with the AFM tip beyond the rupture force in a force vs. z-piezo displacement experiment (Taken from Ref. [69]. Reproduced with permission)

resonance while a second feedback (working over the phase lock loop) changes the tip-sample gap to keep a set point frequency, such that its output gives the topography. Lastly, a third feedback is used to maintain the oscillation amplitude constant by changing the amplitude of the cantilever driving signal, which results in more stable operation.

Finally, it is important to stress the ability of AFM to mechanically manipulate and modify individual virus particles. The first experimental approaches, published very recently, are based upon the application of mechanical stress above the rupture force, thus provoking a rather uncontrolled dismantling of virus particles (Fig. 8.12) [53, 54, 69]. In the future the adequate control of the mechanical stress dosage may unveil further important insights into virus assembly, disassembly and dynamics [1].

In summary, the ongoing developments in AFM promise to improve its applicability in structural virology by: (i) further reducing unwanted sample damage; (ii) increasing the spatial resolution; (iii) increasing the temporal resolution so that relatively fast structural changes in virus particles can be followed in real time.

Acknowledgements I thank my students Aida Llauró-Portell, Alvaro Ortega-Esteban and Mercedes Hernando-Pérez, who are carrying out the hardest part of the work. I also want to thank my collaborators Nuria Verdaguer, Mauricio G. Mateu, José López Carrascosa, David Reguera, Julio Gómez-Herrero and Carmen San Martín. I acknowledge funding by grants from the Ministry of Science and Innovation of Spain, PIB2010US-00233, FIS2011-29493, Consolider CSD2010-00024, Comunidad de Madrid No. S2009/MAT-1467, and FIS2011-16090-E.

References and Further Reading

1. Ortega-Esteban A, Pérez-Berná AJ, Menéndez-Conejero R, Flint SJ, San Martín C, de Pablo PJ (2013) Monitoring dynamics of human adenovirus disassembly induced by mechanical fatigue. Sci Rep 3: 1434
2. Kuznetsov YG, Xiao CA, Sun SY, Raoult D, Rossmann M, McPherson A (2010) Atomic force microscopy investigation of the giant mimivirus. Virology 404:127–137
3. Roos WH, Radtke K, Kniesmeijer E, Geertsema H, Sodeik B, Wuite GJL (2009) Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids. Proc Natl Acad Sci USA 106:9673–9678
4. Kuznetsov YG, Low A, Fan H, McPherson A (2004) Atomic force microscopy investigation of wild-type Moloney murine leukemia virus particles and virus particles lacking the envelope protein. Virology 323:189–196
5. Carrasco C, Castellanos M, de Pablo PJ, Mateu MG (2008) Manipulation of the mechanical properties of a virus by protein engineering. Proc Natl Acad Sci USA 105:4150–4155
6. Hernando-Pérez M, Pascual E, Carrasco C, Ionel A, Carrascosa JL, de Pablo PJ (2009) Study of the mechanical properties of bacteriophage T7. Biophys J 96(3):422a–423a
7. Binnig G, Rohrer H (1982) Scanning tunneling microscopy. Helv Physica Acta 55:726–735
8. Chen CJ (1993) Introduction to scanning tunneling microscopy. Oxford University Press, Oxford
9. Baro AM, Miranda R et al (1985) Determination of surface topography of biological specimens at high-resolution by scanning tunnelling microscopy. Nature 315:253–254
10. Binnig G, Quate CF, Gerber C (1986) Atomic force microscope. Phys Rev Lett 56:930–933
11. Schmalz G (1929) Uber Glatte und Ebenheit als physikalisches und physiologishes Problem. Verein Deutscher Ingenieure (Oct 12), pp. 1461–1467
12. Meyer G, Amer NM (1988) Novel optical approach to atomic force microscopy. Appl Phys Lett 53:1045–1047
13. Sader JE, Chon JWM, Mulvaney P (1999) Calibration of rectangular atomic force microscope cantilevers. Rev Sci Instrum 70:3967–3969
14. Butt HJ, Jaschke M (1995) Calculation of thermal noise in atomic-force microscopy. Nanotechnology 6:1–7
15. Israelachvili J (2002) Intermolecular and surface forces. Academic Press, London
16. Johnson KL (1985) Contact mechanics. Cambridge University Press, Cambridge
17. Ohnesorge F, Binnig G (1993) True atomic-resolution by atomic force microscopy through repulsive and attractive forces. Science 260:1451–1456
18. Giessibl FJ (1995) Atomic-resolution of the silicon (111)-(7x7) surface by atomic-force microscopy. Science 267:68–71
19. Sugimoto Y, Pou P, Abe M, Jelinek P, Perez R, Morita S, Custance O (2007) Chemical identification of individual surface atoms by atomic force microscopy. Nature 446:64–67
20. Carpick RW, Ogletree DF, Salmeron M (1997) Lateral stiffness: a new nanomechanical measurement for the determination of shear strengths with friction force microscopy. Appl Phys Lett 70:1548–1550
21. Martin Y, Williams CC, Wickramasinghe HK (1987) Atomic force microscope force mapping and profiling on a sub 100-a scale. J Appl Phys 61:4723–4729
22. Garcia R, Perez R (2002) Dynamic atomic force microscopy methods. Surf Sci Rep 47:197–301
23. Legleiter J, Park M, Cusick B, Kowalewski T (2006) Scanning probe acceleration microscopy (SPAM) in fluids: mapping mechanical properties of surfaces at the nanoscale. Proc Natl Acad Sci 103:4813–4818
24. Kuznetsov YG, Malkin AJ, Lucas RW, Plomp M, McPherson A (2001) Imaging of viruses by atomic force microscopy. J Gen Virol 82:2025–2034

25. Miyatani T, Horii M, Rosa A, Fujihira M, Marti O (1997) Mapping of electrical double-layer force between tip and sample surfaces in water with pulsed-force-mode atomic force microscopy. Appl Phys Lett 71:2632–2634
26. de Pablo PJ, Colchero J, Gomez-Herrero J, Baro AM (1998) Jumping mode scanning force microscopy. Appl Phys Lett 73:3300–3302
27. de Pablo PJ, Colchero J, Gomez-Herrero J, Baro AM, Schaefer DM, Howell S, Walsh B, Reifenberger R (1999) Adhesion maps using scanning force microscopy techniques. J Adhes 71:339–356
28. Ortega-Esteban A, Horcas I, Hernando-Pérez M, Ares P, Pérez-Berná AJ, San Martín C, Carrascosa JL, de Pablo PJ, Gómez-Herrero J (2012) Minimizing tip-sample forces in jumping mode atomic force microscopy in liquid. Ultramicroscopy 114:56–61
29. Villarrubia JS (1997) Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. J Res Natl Inst Stan Technol 102:425–454
30. Marti O, Drake B, Gould S, Hansma PK (1988) Atomic resolution atomic force microscopy of graphite and the native oxide on silicon. J Vac Sci Technol a-Vac Surf Films 6:287–290
31. Soler JM, Baro AM, Garcia N, Rohrer H (1986) Interatomic forces in scanning tunneling microscopy—giant corrugations of the graphite surface. Phys Rev Lett 57:444–447
32. Hansma HG, Sinsheimer RL, Li MQ, Hansma PK (1992) Atomic force microscopy of single-stranded and double-stranded DNA. Nucleic Acids Res 20:3585–3590
33. Lyubchenko YL, Jacobs BL, Lindsay SM, Stasiak A (1995) Atomic-force microscopy of nucleoprotein complexes. Scanning Microscopy 9:705–727
34. Dame RT, Wyman C, Goosen N (2003) Insights into the regulation of transcription by scanning force microscopy. Journal of Microscopy-Oxford 212:244–253
35. Janicijevic A, Ristic D, Wyman C (2003) The molecular machines of DNA repair: scanning force microscopy analysis of their architecture. Journal of Microscopy-Oxford 212:264–272
36. Wagner P, Hegner M, Guntherodt HJ, Semenza G (1995) Formation and in-situ modification of monolayers chemisorbed on ultraflat template-stripped gold surfaces. Langmuir 11:3867–3875
37. Muller DJ, Amrein M, Engel A (1997) Adsorption of biological molecules to a solid support for scanning probe microscopy. J Struct Biol 119:172–188
38. Muller DJ, Janovjak H, Lehto T, Kuerschner L, Anderson K (2002) Observing structure, function and assembly of single proteins by AFM. Prog Biophys Mol Biol 79:1–43
39. Hansma HG, Pietrasanta L (1998) Atomic force microscopy and other scanning probe microscopies. Curr Opin Chem Biol 2:579–584
40. Horber JKH, Miles MJ (2003) Scanning probe evolution in biology. Science 302:1002–1005
41. Hinterdorfer P, Dufrene YF (2006) Detection and localization of single molecular recognition events using atomic force microscopy. Nat Meth 3:347–355
42. Neuman KC, Nagy A (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Meth 5:491–505
43. Kirmizis D, Logothetidis S (2010) Atomic force microscopy probing in the measurement of cell mechanics. Int J Nanomed 5:137–145
44. Kurland NE, Drira Z, Yadavalli VK (2012) Measurement of nanomechanical properties of biomolecules using atomic force microscopy. Micron 43:116–128
45. Viani MB, Pietrasanta LI, Thompson JB, Chand A, Gebeshuber IC, Kindt JH, Richter M, Hansma HG, Hansma PK (2000) Probing protein-protein interactions in real time. Nat Struct Biol 7:644–647
46. Moreno-Herrero F, Colchero J, Gomez-Herrero J, Baro AM (2004) Atomic force microscopy contact, tapping, and jumping modes for imaging biological samples in liquids. Phys Rev E 69:031915
47. Xu X, Carrasco C, de Pablo PJ, Gomez-Herrero J, Raman A (2008) Unmasking imaging forces on soft biological samples in liquids when using dynamic atomic force microscopy: a case study on viral capsids. Biophys J 95:2520–2528

48. Kasas S, Thomson NH, Smith BL, Hansma HG, Zhu X, Guthold M, Bustamante C, Kool ET, Kashlev M, Hansma PK (1997) Escherichia coli RNA polymerase activity observed using atomic force microscopy. Biochemistry 36:461–468
49. Moreno-Herrero F, de Jager M, Dekker NH, Kanaar R, Wyman C, Dekker C (2005) Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. Nature 437:440–443
50. Kodera N, Yamamoto D, Ishikawa R, Ando T (2010) Video imaging of walking myosin V by high-speed atomic force microscopy. Nature 468:72–76
51. Plomp M, Rice MK, Wagner EK, McPherson A, Malkin AJ (2002) Rapid visualization at high resolution of pathogens by atomic force microscopy – Structural studies of herpes simplex virus-1. Am J Pathol 160:1959–1966
52. Kuznetsov YG, McPherson A (2011) Atomic force microscopy in imaging of viruses and virus-infected cells. Microbiol Mol Biol Rev 75:268–285
53. Ivanovska IL, Miranda R, Carrascosa JL, Wuite GJL, Schmidt CF (2011) Discrete fracture patterns of virus shells reveal mechanical building blocks. Proc Natl Acad Sci USA 108:12611–12616
54. Castellanos M, Perez R, Carrillo PJP, Pablo PJ, Mateu MG (2012) Mechanical disassembly of single virus particles reveals kinetic intermediates predicted by theory. Biophys J 102:2615–2624
55. Sieben C, Kappel C, Zhu R, Wozniak A, Rankl C, Hinterdorfer P, Grubmüller H, Herrmann A (2012) Influenza virus binds its host cell using multiple dynamic interactions. Proc Natl Acad Sci USA 109:13626–13631
56. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG (2009) Structural studies of the giant mimivirus. PLoS Biol 7:958–966
57. Vinckier A, Heyvaert I, Dhoore A, Mckittrick T, Vanhaesendonck C, Engelborghs Y, Hellemans L (1995) Immobilizing and imaging microtubules by atomic-force microscopy. Ultramicroscopy 57:337–343
58. Carrasco C, Luque A, Hernando-Pérez M, Miranda R, Carrascosa JL, Serena PA, de Ridder M, Raman A, Gómez-Herrero J, Schaap IA, Reguera D, de Pablo PJ (2011) Built-in mechanical stress in viral shells. Biophys J 100:1100–1108
59. Roos WH, Bruinsma R, Wuite GJL (2010) Physical virology. Nat Phys 6:733–743
60. Ivanovska IL, de Pablo PJ, Ibarra B, Sgalari G, MacKintosh FC, Carrascosa JL, Schmidt CF, Wuite GJ (2004) Bacteriophage capsids: Tough nanoshells with complex elastic properties. Proc Natl Acad Sci USA 101:7600–7605
61. Kienberger F, Zhu R, Moser R, Blaas D, Hinterdorfer P (2004) Monitoring RNA release from human rhinovirus by dynamic force microscopy. J Virol 78:3203–3209
62. Tang JH, Olson N, Jardine PJ, Grimes S, Anderson DL, Baker TS (2008) DNA poised for release in bacteriophage phi 29. Structure 16:935–943
63. Horcas I, Fernandez R, Gomez-Rodriguez JM, Colchero J, Gomez-Herrero J, Baro AM (2007) WSXM: A software for scanning probe microscopy and a tool for nanotechnology. Rev Sci Instrum 78:013705
64. Carrasco C, Carreira A, Schaap IAT, Serena PA, Gomez-Herrero J, Mateu MG, de Pablo PJ (2006) DNA-mediated anisotropic mechanical reinforcement of a virus. Proc Natl Acad Sci USA 103:13706–13711
65. Ando T, Kodera N, Takai E, Maruyama D, Saito K, Toda A (2001) A high-speed atomic force microscope for studying biological macromolecules. Proc Natl Acad Sci USA 98:12468–12472
66. Melcher J, Carrasco C, Xu X, Carrascosa JL, Gómez-Herrero J, de Pablo PJ, Raman A (2009) Origins of phase contrast in the atomic forces microscopy in liquids. Proc Nat Acad Sci 106:13655–13660
67. Hoogenboom BW, Hug HJ, Pellmont Y, Martin S, Frederix PLTM, Fotiadis D, Engel A (2006) Quantitative dynamic-mode scanning force microscopy in liquid. Appl Phys Lett 88:193109

68. Martinez-Martin D, Carrasco C, Hernando-Perez M, de Pablo PJ, Gomez-Herrero J, Perez R, Mateu MG, Carrascosa JL, Kiracofe D, Melcher J, Raman A (2012) Resolving structure and mechanical properties at the nanoscale of viruses with frequency modulation atomic force microscopy. PLoS One 7:e30204
69. Pérez-Berná AJ, Ortega-Esteban A, Menéndez-Conejero R, Winkler DC, Menéndez M, Steven AC, Flint SJ, de Pablo PJ, San Martín C (2012) The role of capsid maturation on adenovirus priming for sequential uncoating. J Biol Chem 287:31582–31595

Further Reading

Samorí P (ed) (2006) Scanning force microscopies beyond imaging. Wiley-VCH Weinheim, Germany
Baró A, Reifenberger R (2012) Atomic force microscopy in liquid. Wiley-VCH Weinheim, Germany

Also especially recommended for further reading are references [3, 4, 7, 21, 22, 45, 49] *listed above*.

Chapter 9
Optical Tweezers to Study Viruses

J. Ricardo Arias-Gonzalez

Abstract A virus is a complex molecular machine that propagates by channeling its genetic information from cell to cell. Unlike macroscopic engines, it operates in a nanoscopic world under continuous thermal agitation. Viruses have developed efficient passive and active strategies to pack and release nucleic acids. Some aspects of the dynamic behavior of viruses and their substrates can be studied using structural and biochemical techniques. Recently, physical techniques have been applied to dynamic studies of viruses in which their intrinsic mechanical activity can be measured directly. Optical tweezers are a technology that can be used to measure the force, torque and strain produced by molecular motors, as a function of time and at the single-molecule level. Thanks to this technique, some bacteriophages are now known to be powerful nanomachines; they exert force in the piconewton range and their motors work in a highly coordinated fashion for packaging the viral nucleic acid genome. Nucleic acids, whose elasticity and condensation behavior are inherently coupled to the viral packaging mechanisms, are also amenable to examination with optical tweezers. In this chapter, we provide a comprehensive analysis of this laser-based tool, its combination with imaging methods and its application to the study of viruses and viral molecules.

Keywords Biophysics • Virus • Bacteriophage • Capsid • DNA • RNA • Molecular motor • Machine • Single-molecule • Mechanochemistry • Optics • Optical tweezers • Magnetic tweezers • Dynamics • Manipulation • Force • Pressure • Elasticity • Condensation • DNA packaging

J.R. Arias-Gonzalez (✉)
Instituto Madrileño de Estudios Avanzados en Nanociencia (IMDEA Nanociencia), c/Faraday 9, Campus de Cantoblanco, 28049 Madrid, Spain

Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain

CNB-CSIC-IMDEA Nanociencia associated unit "Unidad de Nanobiotecnología", Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: ricardo.arias@imdea.org

M.G. Mateu (ed.), *Structure and Physics of Viruses; An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_9,

Abbreviations

3D	Three dimensional
AFM	Atomic force microscopy
bp	Base pair
Dig	Digoxigenin
ds	Double-stranded
GC	Guanine-cytosine
IR	Infrared
NA	Numerical aperture
α-Dig	Anti-digoxigenin

9.1 Introduction: Life Machinery at the Nanoscale

Optical tweezers are an instrument able to measure the forces and motion involved in the dynamics of biomolecular motors. From this viewpoint, they constitute a unique instrument to study the physics underlying some viral processes. In contrast to other structural techniques described in Chaps. 3, 4, 5, 7 and 8 that allow the visualization of virions and capsids, dynamic experiments using optical tweezers are unable to 'photograph' the main characters of this book, but can quantitatively track their mechanical activity in real time. This chapter will introduce a new method with which the virus is followed while it is at work, packaging its nucleic acid into the viral capsid. Characteristic results of these experiments are forces and packaging velocities as a function of time or as a function of the amount of packaged genome. A related problem is the mechanical behavior of nucleic acids; force-extension curves of DNA/RNA to determine their elasticity can be also obtained with this technique.

The machinery that makes up life is very small and highly condensed. The dimensions of a globular protein or the thickness of DNA are measured in nanometers. This machinery works in crowded environments (in a cell, the total concentration of macromolecules may reach 300–400 g/l), and is subject to continuous agitation. In these conditions, movement is strongly affected by thermal fluctuations (Brownian motion), and structures are markedly distorted by mechanical stress. The dynamic picture of the motion and activity of a protein or macromolecular complex is that of a boat sailing against furious winds and waves, with rapid changes in direction. Throughout evolution, proteins have developed strategies to rectify and bias thermal fluctuations using the energy from ATP or GTP in a way similar to the orientation of the boat's sail and rudder. These strategies are different from the macroscopic, human-made machinery, whose dynamic behavior more closely resembles a motorboat in a quiet sea, for which fuel energy is used to propel the boat in a chosen direction. Biological machinery is stochastic and soft, unlike macroscopic machines, whose movements are predictable and whose structures

are typically rigid. Biological machinery is nonetheless highly efficient in terms of the ratio [developed work]/[spent energy], much more so than human technology to date. A further singularity of biomolecular machinery is its ability to interact with its environment. Biological machines are in essence biochemical entities; they change shape, size and mass as a result of their function. The action of most proteins and macromolecular assemblies, including viruses, is thus *mechanochemical*, and their complete study must include not only biochemical aspects, but also mechanical analyses (see also Chaps. 8 and 18).

Optical tweezers, described in this chapter, is a technique that allows such mechanochemical analysis. Viruses, like other large biological structures such as ribosomes, ATP synthase or the bacterial flagellum, are complex machines, and not passive structures. Many complex viruses in particular are coordinated assemblies of motors, structural scaffolds and information carriers. Many bacteriophages are fascinating from a technological point of view, because they are some of the most powerful machines at the nanoscale acting as packaging devices for their own nucleic acid. In turn, the nucleic acids represent soft 'hard drives' that store and spread information. Some nucleic acid physical properties, such as their elasticity and denaturation behavior, may have evolved in tune with the machinery that processes them, rendering these processes more efficient over the ages. Understanding DNA and RNA from a mechanochemical point of view may thus shed light on viral performance. This chapter will also be a platform to introduce these novel views of the viral machinery and its peripherals.

9.2 Basic Concepts and General Experimental Design

The forces developed by protein machines including viruses are in the piconewton (pN) range, which compares with one trillionth (10^{-12}) the weight of a lemon. Measuring these tiny forces was not technologically easy before the invention of the laser in the 1960s; use of this light technology to move and trap micro- and nanostructures began in the 1970s. By the mid-1980s, this application had been developed into a technique commonly known as optical or laser tweezers. Its use in molecular and cell biology became relevant by the 1990s, although experiments manipulating viruses, cells or organelles within living cells already saw the light in the previous decade [1].

The forces generated by electromagnetic radiation over a physical object are termed *optical forces*. The strength of these forces is generally in the piconewton range, although this depends on the material properties of the object, on the radiation power, and on beam shape. An instrument based on the use of optical forces will therefore be sensitive to the tiny forces exerted by molecular machinery. Since electromagnetic radiation can penetrate the medium without traceable effects, optical forces are a non-contact solution to measuring and exerting force

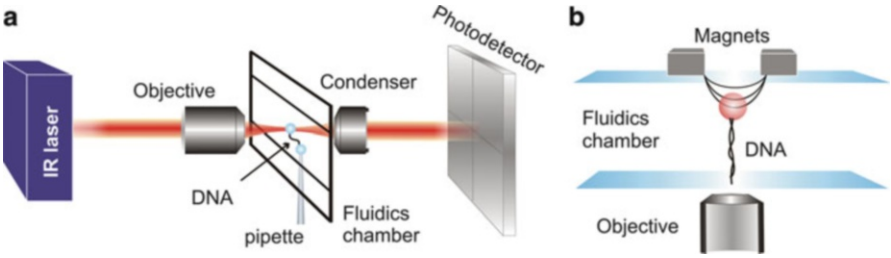


Fig. 9.1 Optical and magnetic tweezers. (**a**) Laser light is highly focused by a microscope objective and collected by a condenser lens. A bead or cell can be trapped in the focal region, which is produced inside a fluidics chamber. A micropipette inserted in the chamber can hold a second bead by suction. Micrometer-sized beads are typically used as mechanical handles to hold a biopolymer by both ends. The collected exiting rays are registered by a position-sensitive photodetector. The scheme shows a single-molecule experiment in which a DNA fiber is stretched between two beads. (**b**) The same experiment is performed using two small magnets that pull a magnetic microsphere inside a fluidics chamber. The DNA is held between the microsphere and the bottom coverslip. A microscope objective is used to monitor the experiment

on biological particles in their liquid environment and within living cells. Inertial and gravitational effects are insignificant in the timescales associated with the dynamics of most biological systems.

For reasons that will become clear later, a strongly focused laser beam generates optical forces in all directions towards the focal region, provided that these objects do not absorb light. The result is a three-dimensional (3D) *optical trap*. Most biological structures, natural or synthetic, do not substantially absorb visible or infrared (IR) radiation and can therefore be maneuvered by a strongly focused laser beam within these spectral bandwidths. If the light scattered from the optical trap is analyzed, the forces acting on the trapped object can be deduced. This is the basis of the optical tweezers, an instrument able to trap nanometer and micrometer-sized particles and to measure the external forces that act on them. The resulting displacements can be measured in the sub-nanometer range with millisecond time resolution.

A simplified version of an optical tweezers instrument is shown in Fig. 9.1 (in comparison to its magnetic counterpart, described in Sect. 9.3.7). The basic elements in the optical tweezers can be grouped into those common to an optical microscope and those related to the generation of the optical trap [2]. The *microscope* elements are an objective, a fluidics chamber and a condenser lens. The *trap* elements are a laser and a photodetector that registers the light at the exit of the condenser. To produce a strongly focused beam, a high-numerical aperture (NA) objective is required (water or oil immersion). Near-IR lasers are optimal for manipulating biological structures, because neither water nor biological structures strongly absorb radiation at this wavelength. Laser power in the focal region of several hundred milliwatts provides a good balance that prevents photodamage and produces forces up to a few hundreds of piconewtons. Magnetic tweezers use two small magnets to generate high field gradients in the vicinity of a paramagnetic

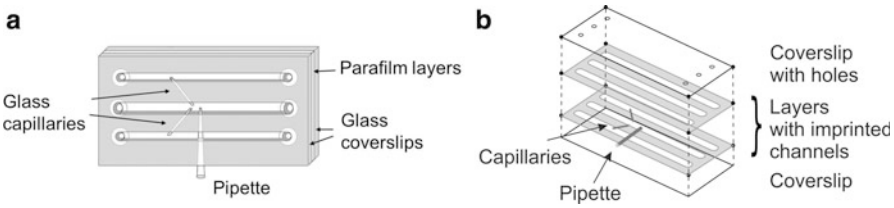


Fig. 9.2 Fluidics chamber. (**a**) The specimens (cells or beads) are passed sequentially through a system of channels. In this design, the experiment takes place in the central channel, which is connected to the *top* and the *bottom* channels by glass capillaries. The scheme also shows the disposition of a micropipette, useful in a variety of experiments. (**b**) Mounting a custom-made fluidics chamber for optical tweezers. After assembling the tubes and parafilm layers, the components are heat-sealed

bead, as shown in Fig. 9.1b. An objective behind the fluidics chamber is used to follow the experiment. This instrument is a good alternative for low forces and to generate torque, as we will see in Sect. 9.3.7.

The fluidics chamber is the test tube for mechanochemical assays. It is usually built using two microscope coverslips sandwiching parafilm layers with imprinted channels. The design shown in Fig. 9.2 allows specimens to be flowed through a top and a bottom channel connected with the central channel by glass capillaries. The optical trap is attained in the central channel. Buffers can be easily exchanged and flowed at a controlled rate, thus adding the ability to exert hydrodynamic pressure. A micropipette can be placed inside the chamber to suck beads and additional tubes can be used to dispense chemicals (for example, ATP) directly to the vicinity of the optical trap. These alternatives are very convenient in a variety of experiments that will be described later in this chapter. In contrast to most structural techniques, optical tweezers allow experiments in physiological conditions that can be changed easily and rapidly (for example, chemical concentration and pH) without affecting the performance of the instrument, which allow the detection in real time of processes triggered by such changes in conditions.

9.2.1 Single-Molecule Approach

As indicated by the above description, this method allows one-by-one trapping and manipulation of specimens (individual molecules and macromolecular assemblies, including virus particles). This ability is the basis of the *single-molecule* approach [3]. *Bulk* experiments, in which a large number of identical (but unavoidably unsynchronized) reactions take place, hinder access to signals from intermediate states of the specimens. Ensemble methods such as electron microscopy, X-ray crystallography or nuclear magnetic resonance (NMR) allow 'taking snapshots' of specific instants in the temporal flow of events for a certain type of molecule, but only the average state is eventually reconstructed from the data.

In addition, data and reconstructed states might be influenced by the cooperative disposition (periodic or non-periodic) of the ensemble before the experiment, yielding misleading structural information or masking rare but potentially important conformations. In fact, bulk experiments only allow observation of average dynamics. Single-molecule experiments permit the unraveling of real-time dynamics and conformations, because the signals are not confused by multiple, asynchronous events.

Another drawback of bulk experiments is that they tend to smear out the information from *thermal fluctuations*. The idealized dynamic picture envisaged from these experiments is a 'movie' at long timescales, in which molecular trajectories are smooth and well defined. At short timescales, this movie is very different: dynamics are fast and random (stochastic). As explained in the introduction, proteins and viruses dwell in a background of thermal agitation that affects their dynamic performance. This inherent dynamics can only be captured by real-time monitoring of individual processes. Fluctuations represent an integral part of protein function, and single-molecule information must therefore be included in the overall description of biomolecular processes. The average energy of a thermal fluctuation is k_BT, where k_B is the Boltzmann constant and T, the temperature. At 25 °C, this energy is ~$4.1 \cdot 10^{-21}$ J, a notable number if we consider the free energy released in the hydrolysis of one ATP molecule at physiological conditions, ~20–24 k_BT.

9.2.2 Molecular Machinery

Living in a thermal bath of violent agitations might seem to render any machinery inefficient. As pointed out in the introduction, however, biological machines have adapted to this apparently uncomfortable situation by developing a mechanism to extract energy from the thermal bath. This mechanism, the *Brownian ratchet*, consists roughly of rectifying random pushes (like a sailboat) using energy from ATP/GTP. In cases in which thermal pushes are not sufficient (for example, in large structures like the bacterial flagellum), protein machinery might bias diffusional steps by burning ATP as a *power stroke* (like a motorboat), a mechanism similar to that of macroscopic engines. In general, both mechanisms are combined by proteins in an intricate fashion to generate mechanical work [4].

9.2.3 Limitations of Optical Tweezers

Optical tweezers do not perform real-time microscopy, *i.e.*, structures cannot be resolved during the assay. Optical tweezers provide measurements of force, distance and time, but cannot ascertain the disposition of protein domains at a given moment or the structure of nucleic acids, for example, under a stretching force. The

specimens are grossly considered as point-like or fiber-like structures. Combining fluorescence with optical tweezers (see perspectives in Sect. 9.6) can help determine the position and movement of a certain molecule, but in most cases the experimenter is 'blind': only the position of micrometer-sized structures, such as cells or the beads shown in Fig. 9.1, can be observed during the assay. In these conditions, structural and biochemical information must be known in advance if experiments are to be designed precisely and correctly interpreted.

9.3 Optical Tweezers

9.3.1 Optical Forces

Light can be viewed for many purposes as a flux of small particles called photons, which can interact with other objects and exchange energy as well as linear and angular momentum, thus generating forces and torques. This simple reasoning is the physical basis underlying optical forces: due to light-matter interaction, light incident on an object can push or modify the object's trajectory in complex ways. These forces are very small, however, and we therefore do not feel the so-called radiation pressure from the Sun, for example, or from the light of a bulb. The fact that light interacts with matter nonetheless explains photodamage and mutations by ultraviolet or X-ray radiation at long exposure times and powers. These radiations are made up of very high energy photons which, rather than pushing biological matter, alter its structure at the atomic level. At visible and IR wavelengths, radiation is made up of photons with much less energy; in this case, the stream of coherent photons pushes small objects in a similar fashion as water exiting a hose at high pressure. This radiation pressure explains, for instance, why a comet's dust tail points away from the Sun. Radiation pressure was long ago suggested as a propulsion method for spacecraft, by using thin mirrors as 'solar sails', although such sails would require areas of square kilometers.

Optical forces are therefore too tiny to be practical for the manipulation of macroscopic objects, but their effect is patent on atoms and on micrometer- and nanometer-sized objects. By adequately shaping the photon bundle from a laser source, it is possible not only to push tiny structures but also to trap them in 3D. The first optical trap was designed to manipulate atoms, which paved the way for the study of many interesting physical phenomena [5]. Trapping of small particles that scale to a few nanometers came soon afterwards, an interest that remains active for nanotechnological applications. Trapping of cells was then tested. Whereas light manipulation of small objects and atoms is useful for understanding light-matter interactions, however, the interest of optical traps in biochemical research resides in their use as force sensors, and as a method to mediate mechanochemical reactions.

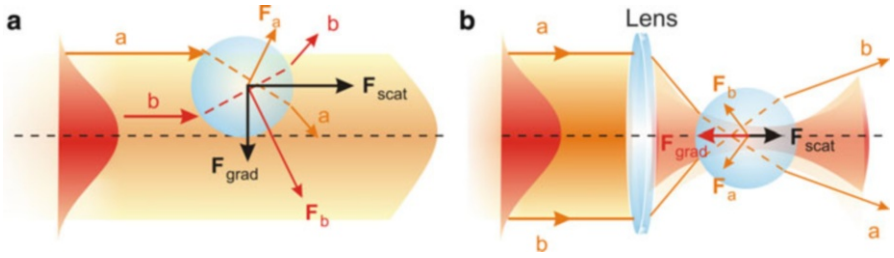


Fig. 9.3 Optical forces acting on a non-absorbing sphere. (**a**) Qualitative view of the forces exerted by a Gaussian laser beam on a sphere. The diagram illustrates the origin of the scattering (F_{scat}) and gradient (F_{grad}) components of the optical force. (**b**) Qualitative view of the forces exerted by a focused Gaussian beam on a sphere on-axis. The diagram shows the axial stability established by the gradient force along the centerline of the beam, towards the focal region. See text for details

9.3.2 An Optical Trap

Here we will analyze how to engineer the light from a laser source to confine an object in 3D. The internal structure of a complex biological particle cannot be resolved by visible light wavelengths; consequently, structure is not essential to understand the trapping analysis. We will assume that the particle is dielectric (*i.e.*, non-absorbing) and has a low refractive index (~1.5). These assumptions are common to biological structures including viruses and many synthetic particles. The main effect of light on this kind of object is refraction. The physics of trapping absorbing objects (metallic or semiconductor particles) or non-absorbing particles under resonant conditions are more complex, but to date they have not been exploited in biological experiments. In addition, absorbing and resonant light effects in biological structures are not desired if one wants to observe biological function without perturbation. These effects will therefore not be discussed here.

Figure 9.3a shows a spherical particle under the action of a laser beam whose wavelength is smaller than the size of the sphere. A ray of light can be conceived as a stream of photons, each carrying an associated momentum. The photons following the trajectories labeled "*a*" and "*b*" will be refracted (neglecting relatively minor surface reflections), as depicted, due to their interaction with the particle; as a result, their momentum will change. The momentum change experienced by each photon is transferred to the particle, thus generating the forces F_a and F_b. As more photons follow path "*b*" than "*a*" – because beam intensity is higher along the former path – force F_b is stronger than F_a. After decomposition of vectors F_a and F_b in two axes, one in the propagation direction of the laser and the other perpendicular to it, and extension of this ray analysis to all symmetrical pairs of rays, the net effect of laser light can be understood in terms of two forces: the *scattering force*, which pushes the particle along the incident beam, and the *gradient force*, which points towards the centerline of the beam. From the diagram, it is clear that there is transverse stability for a particle on-axis, since here $F_a = F_b$. To establish a 3D

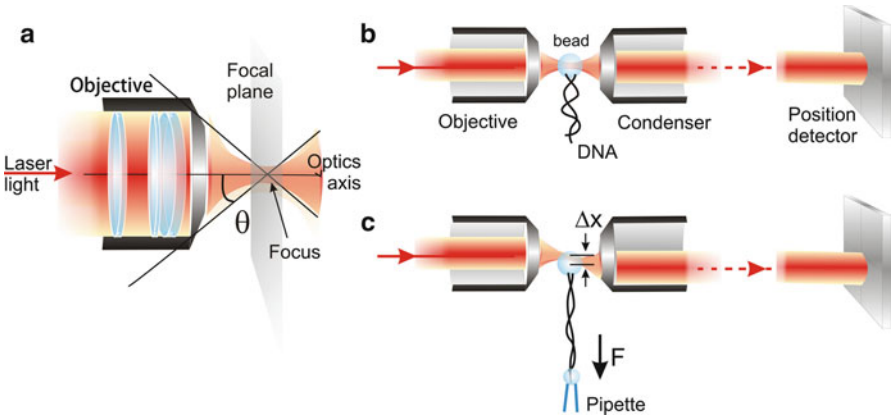


Fig. 9.4 Trapping and detecting force and displacements. (**a**) Representation of a microscope objective and the aperture angle, θ, of the cone of rays which are made to converge to a focus. In this configuration, laser light overfills the objective. (**b**) Diagram illustrating the configuration of the objective in front of a condenser lens, used to collect the light at the exit of the trap, and a photodetector to analyze the scattered light. Between the condenser lens and the photodetector, there are other optical elements not shown here for simplicity (see Fig. 9.5). (**c**) Diagram showing how an external force can be measured by registering light flux offsets in the photodetector. In (**b**) and (**c**), laser light underfills the objectives. Diagrams not to scale

trap, axial stability must be achieved, that is, the scattering force must be compensated as well. Arthur Ashkin first made this possible using two equal laser beams facing each other; a trap generated in this fashion is known as a *dual-beam trap*. Some years later, by focusing the laser beam with an objective lens, he showed that a single-beam trap is also possible (find advanced literature in [1]). Figure 9.3b shows the mechanical stability of a particle in such a trap: the focal region is now a global maximum of optical intensity and in these conditions, an axial gradient force pointing towards the focus is generated before and after the light passes through that region. If the focus is sufficiently strong, the gradient effect also dominates the axial stability thus resulting in the all-optical capture of the particle.

9.3.3 Optical Trapping and Manipulation of Micro-and Nanoparticles

The photon flux from a laser source describes approximately parallel, linear trajectories at the output. Using simple optical elements like lenses and mirrors, the photon flux can be guided to a microscope objective (Fig. 9.4a). The objective can be viewed as a lens that causes the parallel trajectories to converge on an ideal point called the focus. In this approximate description, we do not consider that photons traveling far from the center of the bundle interact with the edges of the objective. These photons will deviate from the linear, perfectly converging

trajectories, hence will also slightly affect the trajectories of the remainder of the photons in the bundle. These effects are known more rigorously as *diffraction*; the result is that the photon bundle does not converge to a point, but to a focal volume with a size on the order of the light wavelength.

Laser light in fact behaves not only as a set of traveling particles, but also like a wave (*wave-particle duality*). It is actually an electromagnetic wave, whose electric and magnetic fields oscillate in both space and time. Laser light is characterized by its wavelength, which is a measure of the distance along which the electromagnetic wave describes a complete oscillation, and by its power in units of energy per transversal area. The intensity of the electromagnetic flux at the output of the laser is highest for the central photon trajectory and decreases rapidly for trajectories away from the center. This type of "light bundle" is called a Gaussian beam (Fig. 9.3a) and is common to most commercial lasers. Other names for these light bundles are TEM_{00} *beams*, and another term for the laser sources that produce them are *single mode lasers*. For optical trapping, it is common to use continuous wave (*CW*) lasers rather than pulsed lasers; in simple terms, they are switched on/off at high frequencies. The laser powers required for optical tweezers are of some hundreds of mW.

The microscope objective is actually a set of many lenses arranged to produce a high focusing effect (Fig. 9.4a). The focusing power is characterized by the numerical aperture (NA); for reasons that will be explained later, the higher the NA, the higher the trapping strength. Objectives are fabricated to work in air, water or oil, and the higher the refractive index of the immersion medium, the higher the NA of the objective. For optical tweezers applications, it is thus common to use water and oil immersion objectives. Let n be the refractive index of the immersion medium and θ, the angle of convergence of the rays after the objective (Fig. 9.4a). The NA is then defined by the equation $NA = nsin\theta$. The maximum numerical aperture, $NA = n$, is achieved when the rays are ideally made to converge on the focus at 90° angles (in practice, objectives cannot curve the rays that way). Water has a refractive index of 1.333 and usual oils, 1.51. Water immersion objectives normally have $NA = 1.2$ and oil immersion objectives can achieve $NA = 1.4$. As explained briefly in Sect. 9.2, the optical trap is attained in a fluidics chamber (Fig. 9.2), a test tube that will be described later to explain how to perform single-molecule experiments.

9.3.4 Force Measurement

If we only need to use the optical trap to manipulate objects, the key instrumentation described so far is sufficient. If we want to measure forces, however, we need to analyze the light scattered from the optical trap. The light exiting the trap is a diverging light ray bundle that can be collected by another objective or by a cost-effective condenser lens. After the condenser lens, the light is redirected to a position-sensitive photodetector. This type of photodetector measures power and registers

light flux offsets, as depicted in Fig. 9.4b and c. An external force acting on the optically trapped particle deflects the exiting light cone and subsequently changes its projection after the condenser lens on the photodetector plane. Another method widely used in optical tweezers to measure bead displacement is back-focal plane interferometry, in which displacements are deduced from the interference pattern between the trapping laser and the scattered light (find advanced literature in [7]).

The use of modified position-sensitive photodetectors to register laser light intensity distribution is the basis for determining the externally applied force, through the momentum conservation principle [6]. This force calibration method has several advantages (it is independent of particle size, shape or refractive index, of buffer viscosity or refractive index, and of variations in laser power), although a simpler method is used in most optical tweezers design. This method is based on the fact that the external force, F, acting on the optically trapped objects is approximately proportional to the generated displacement, Δx (see Fig. 9.4c). In other words, $F = \kappa \Delta x$, where κ is the trap stiffness, a parameter that characterizes the 'invisible' spring that binds the particle to the optical trap. This parameter depends on the laser power and on the properties of both the particle and the immersion medium. Methods like the use of the Stokes law, or the Brownian motion of the bead in the trap can be used to accurately determine the trap stiffness, as described in-depth in other technical reviews and articles [2].

9.3.5 Optical Tweezers Instrument

The fact that the main elements of an optical microscope are involved in the design of a laser trap made the use of conventional microscopes the first and easiest strategy to set up an optical tweezers instrument. In Sect. 9.2, we divided the main components of the optical tweezers into those corresponding to the 'microscope' and to the 'trap' parts. The different elements are nonetheless optimally tuned by setting all components on an optical table with vibration isolation. Most optical tweezers instruments are homemade designed, with little use of purpose-designed commercial apparatus to date.

Experiments in the optical tweezers use microsphere position as the visual reference. Videomicroscopy can easily be integrated in optical trap design, because the objective that focuses the laser beam can also be used to generate real-time images of the events in the focal plane (Fig. 9.4a). A white-light lamp positioned on the condenser lens side is often used to illuminate the fluidics chamber interior in Köhler conditions, as in an optical microscope. A CCD camera at the back focal plane of the objective is used to visualize the experiment. Images from the camera and signals from the photodetectors are processed by a computer.

Figure 9.5 represents the component layout in an optical tweezers instrument. This scheme is not standardized, but integrates the general concepts and components of the instrument. In-depth design considerations are discussed in advanced reviews and technical articles [2, 6, 7]. In this scheme, two counter-propagating lasers

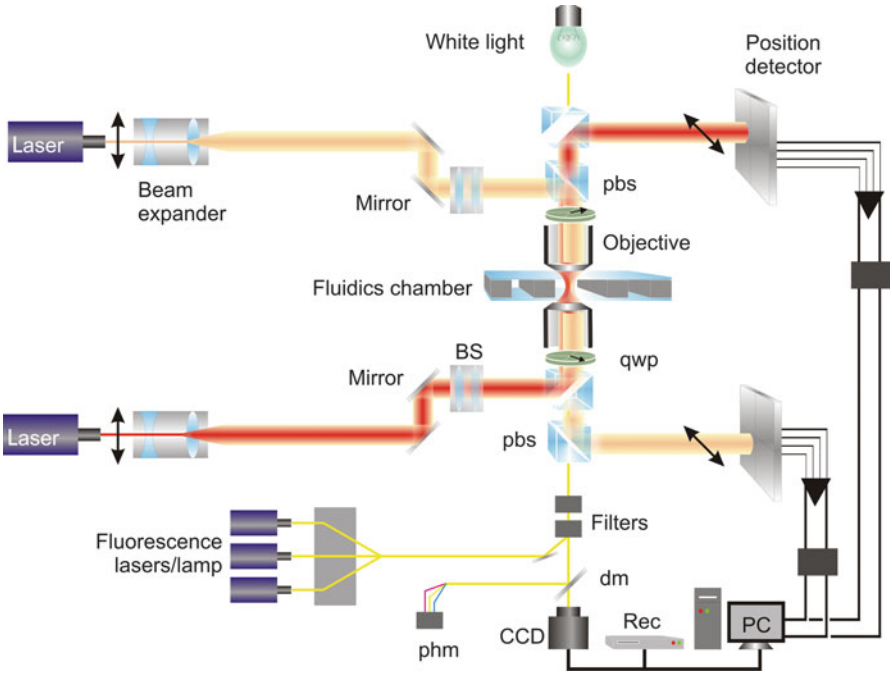


Fig. 9.5 Optical tweezers setup. Schematic summary of the most common designs (not to scale). The *arrows* show polarization of the beams at the exit of the laser sources and before the photodetectors. In this design, the polarizations of the beams in the trap are made circular and mutually orthogonal [6]. Additional light sources can be introduced in the setup to implement confocal microscopy and fluorescence. The laser wavelengths of the confocal module must be different from the infrared wavelengths used to generate the optical trap(s). *BS* beam steering, *pbs* polarizing beam splitter, *qwp* quarter wave plate, *dm* dichroic mirror, *Rec* video recorder, *phm* photomultiplier

converge on the fluidics chamber, for which two objectives are used instead of a pair objective-condenser lenses. If the two lasers converge on the same focus, the instrument is a *dual-beam optical tweezers*. In this design, a micropipette is used to grab a second bead, as in the single-beam optical tweezers described above. If the two lasers are brought to separate focuses, each could be used to grab an independent bead without the need for a micropipette; the instrument so designed is a *dual-trap optical tweezers*. The equipment can also be engineered to split the light from a single laser source into two optical paths, a strategy that increases resolution [8].

In contrast to single-beam optical tweezers, the lasers in dual-beam instruments do not need to overfill the objectives to generate an efficient trap (see Fig. 9.4). Underfilling reduces energy concentration in the focal region, which in turn decreases radiative effects on the trapped specimen. This situation is also exploited to calibrate the optical tweezers by direct measurement of light momentum, as indicated earlier.

When two beams share the same trajectory, it is of paramount importance to avoid interferences. Electromagnetic waves can interfere, generating spurious effects in the optical trap that could lead to undesired consequences in manipulation and measurement. To prevent this, beams from distinct sources should encounter each other with orthogonal polarizations. Polarization is a property of light waves that describes the direction of oscillation of the electric and magnetic fields. Light from a laser source is often linearly polarized, as shown in Fig. 9.5. Quarter-wave plates can rotate the polarization of each laser beam immediately before it enters the microscope objectives, so that they do not interfere with each other in the focal region. Polarizing beam splitters are optical elements that transmit or reflect the light depending on its polarization. These elements can be used to guide the two beams before and after the objectives (Fig. 9.5).

The two beams can be steered using motorized elements, such as mirrors and lenses, which can be computer-controlled. These elements allow moving one trap relative to the other in dual-trap optical tweezers. Beam expanders are also used to adjust the area of the laser spot at the exit of the laser sources to the entrance aperture of the objectives. They are used to over- or underfill the objectives, depending on the optical tweezers design. Beam expanders are normally made up of two lenses of different focal length. A pinhole can be positioned between them to filter the beam spatially. Astigmatism lenses can be introduced in the laser paths to correct this aberration (not shown in Fig. 9.5).

Finally, it is important to say a word about imaging possibilities. Unless fluorescence imaging is used, only the position of microscopic objects will allow monitoring experiments in the optical tweezers. Cells can be tracked easily, as can large organelles like centrosomes [9], but molecules cannot. Experiments with proteins and nucleic acids thus involve their attachment to microspheres in the so-called molecular construct (see Sect. 9.4.2). These microspheres, generally made of polystyrene or silica, act as mechanical handles to tether the specimen at specific sites. Correct design of the molecular construct and as much structural information as possible is necessary to perform and interpret the assays correctly. Single-molecule fluorescence, including the use of quantum dots,[1] and confocal microscopy[2] can be integrated in the optical tweezers; these are of great help to provide dynamic information on the mechanochemical experiments (see Sect. 9.6.3). Figure 9.5 illustrates one way of combining fluorescence in the optical tweezers setup.

[1] *Quantum dots* are luminescent semiconductor nanocrystals which are used as a new class of fluorescent probes. In comparison with organic dyes and fluorescent proteins, they are brighter and exhibit both greater photostability and longer lifetimes. They can also be tuned to specific wavelengths with narrower bandwidths than fluorescent molecules.

[2] *Confocal microscopy* is a high-resolution imaging technique based on a point-by-point illumination of the sample. The light source is often a focused laser beam. Lasers of several wavelengths and filters can be used to combine this technique with fluorescent imaging.

9.3.6 Electromagnetic Theory

The forces that generate a laser source are electromagnetic. In this section, we present a rigorous theory to explain the origin of these interactions on non-absorbing nanoparticles. This section will help to better understand the concepts presented on optical forces in Sect. 9.3.2. The following is not essential for performing experiments or understanding the instrumental description above, and can therefore be skipped.

A particle is small when its characteristic diameter is much smaller than the illumination wavelength. Laser wavelengths for optical tweezers purposes are in the near-IR range, usually from 800 to 1,100 nm. Longer wavelengths are strongly absorbed by water and generate undesired heating and convection effects. A particle is then small (also known as a Rayleigh or dipolar particle) when its diameter is below ~100 nm. The Lorentz force in the presence of an electromagnetic field is [10]:

$$\vec{f} = (\vec{\wp} \cdot \nabla)\vec{\varepsilon} + \frac{\partial \vec{\wp}}{\partial t} \times \vec{\beta} \tag{9.1}$$

where $\vec{\wp}$ is the induced electric dipole moment and $\vec{\varepsilon}, \vec{\beta}$ are the electric and magnetic vectors, respectively (SI units). Experiments in optical tweezers are performed in liquid media and particle motion is therefore greatly slowed by viscous forces. In these conditions, the particles cannot follow the rapid oscillations of the electromagnetic fields; consequently, the time-average force is the observed value. The electromagnetic field from a single mode laser is harmonic and can thus be expressed in complex representation as $\vec{\varepsilon}(\vec{r},t) = \Re e\{\vec{E}(\vec{r})exp(-i\omega t)\}$ and $\vec{\beta}(\vec{r},t) = \Re e\{\vec{B}(\vec{r})\exp(-i\omega t)\}$, and the dipolar moment as $\vec{\wp}(\vec{r},t) = \Re e\{\vec{p}(\vec{r})\exp(-i\omega t)\}$; $\vec{E}, \vec{B}$ y $\vec{p}$ are, in this formalism, complex functions of position. $\Re e$ denotes the real part and ω is the angular frequency of the field. The i-th component of the time-average force is thus (see [11] and references therein):

$$\langle f_i(\vec{r},t)\rangle = F_i(\vec{r}) = \frac{1}{2}\Re e\left\{\alpha \sum_{j=x,y,z} E_j(\vec{r})\frac{\partial E_j^*(\vec{r})}{\partial x_i}\right\}, i = x,y,z \tag{9.2}$$

where * denotes the complex conjugate. Equation (9.2) is the expression of the observed optical force, $\vec{F} = (F_x, F_y, F_z)$, over a small particle with complex electric polarizability α in the liquid medium.

The electric field can be expressed in paraxial representation, that is, a beam or evanescent field with a main propagating direction along $\vec{k}$, as $\vec{E}(\vec{r}) = \vec{E}_0(\vec{r})$ $\exp\left(i\vec{k}\,.\,\vec{r}\right)$. The modulus of vector $\vec{k}$ is given by$k = n(\omega/c)$, being n the refractive

index of the liquid medium and c the speed of light in vacuum. Substituting this expression into Eq. (9.2), one obtains [11]:

$$\vec{F} = \frac{1}{2}\vec{k}\Im m\{\alpha\}\left|\vec{E}_0\right|^2 + \frac{1}{4}\Re e\{\alpha\}\nabla\left|\vec{E}_0\right|^2 - \frac{1}{2}\Im m\{\alpha\}\Im m\left\{\left(\vec{E}_0\cdot\nabla\right)\vec{E}_0^*\right\}, \tag{9.3}$$

where $\Im m$ denotes the imaginary part. The first term in Eq. (9.3) is proportional to the field intensity, $\left|\vec{E}_0\right|^2$, and points along the propagating direction, $\vec{k}$. This term therefore represents the contribution of the radiation pressure to the scattering force. Since $\Im m\{\alpha\} > 0$, this force pushes the particle in the laser-propagating direction. The second term is proportional to the gradient of the field intensity, $\nabla\left|\vec{E}_0\right|^2$, and it is thus known as the gradient force. Since $\Re e\{\alpha\} > 0$ for non-absorbing particles, this force attracts the particle towards the maximum of the field intensity. This term is responsible for optical trapping, as was also explained qualitatively in Sect. 9.3.2. The last term is not significant in typical experiments but may induce rotation of the trapped particle due to spin angular momentum exchange between the field and the particle (see Sect. 9.6.2).

Optical forces on many viral particles can be calculated by using Eq. (9.3). Beads used as mechanical handles are usually in the micrometer range and therefore beyond this approximation. The above description nonetheless allows understanding the physics of the problem, which was previously treated with a ray optics description in Sect. 9.3.2. The force on these larger particles can be calculated using the following expression [10, 11]:

$$\begin{aligned}\vec{F} = &\frac{1}{2}\Re e\left\{\int_\Sigma d^2r\left[\left(\in\vec{E}(\vec{r},\omega)\cdot\vec{n}\right)\vec{E}^*(\vec{r},\omega) + \left(\mu\vec{H}(\vec{r},\omega)\cdot\vec{n}\right)\vec{H}^*(\vec{r},\omega)\right]\right.\\ &\left.-\frac{1}{2}\left(\in\left|\vec{E}(\vec{r},\omega)\right|^2\right) + \left(\mu|H(\vec{r},\omega)|^2\right)\vec{n}\right\}\end{aligned} \tag{9.4}$$

where Σ is a surface enclosing the particle and $\vec{n}$ is its local outward normal unit vector. $\in$ and μ are the electric and magnetic relative permittivities of the medium, respectively. The complex representation of the H-field, $\vec{\aleph}(\vec{r},t) = \Re e\left\{\vec{H}(\vec{r})\exp(-i\omega t)\right\}$, is related to the complex B-field by $\vec{B}(\vec{r}) = \mu\vec{H}(\vec{r})$. Equation (9.4) uses the time-harmonic expression of Maxwell's stress tensor. Computation of forces from this expression involves numerical calculations in most cases. This might be useful to predict the behavior of special particles in the trap but for most biological experiments, it is often only required to characterize empirically the field-particle interaction by a spring constant (see Sect. 9.3.4).

9.3.7 Magnetic Tweezers

Magnetic tweezers are a cost-effective instrument to perform single-molecule experiments with applications similar to those of optical tweezers [12]. This

instrument can be used to measure and exert forces in mechanochemical reactions, although to date it has been little used to study viruses and viral components, with the exception of their nucleic acids. To provide a broad palette of possibilities for virus study, we present this technique briefly.

Typical magnetic tweezers setups (Fig. 9.1b) use permanent magnets to generate a field gradient inside a fluidics chamber. By moving these magnets along the field gradient, forces on magnetic particles in a fluidics chamber can be increased/decreased; by rotating them, torque can be induced. Electromagnets can also be used to generate stretching forces. In this case, adjusting the current running through the magnetic coils allows change in the force strength. Superparamagnetic beads are used as mechanical handles for biological molecules in these experiments. Due to space constraints, magnets must be placed outside the fluidics chamber, which prevents the highest field gradients from approaching the magnetic beads, thus establishing an upper boundary to the forces generated. Maximum forces on micrometer-sized superparamagnetic particles in these conditions are usually below ~10 pN. Recent developments using microscopic magnetic poles inside the fluidics chamber can greatly increase the forces, to a range similar to that of optical tweezers [13]. Minimum generated forces are in the range of 0.01 pN.

The force generated by two magnets on a superparamagnetic bead is

$$\vec{F} = \frac{1}{2}\nabla\left(\vec{m}\cdot\vec{\beta}\right), \tag{9.5}$$

where $\vec{m}$ is the induced magnetic moment on the microsphere in the presence of the magnetic field, $\vec{\beta}$, arising from the magnets. Force calibration can be performed against the viscous drag or using Brownian fluctuations. A microscope objective and a camera below the liquid chamber are used to follow the experiments, specifically, to deduce the forces and displacements of the bead. Illumination comes from above the fluidics chamber in this design (Fig. 9.1b). This apparatus can also be built on an inverted microscope.

Although this instrument constitutes the magnetic analogue of optical tweezers, it is not based on the generation of a trap. In other words, particles are not grabbed but dragged. To be manipulated, single particles must be tethered. For example, the DNA molecule in the experiment in Fig. 9.1b is always subject to a pulling force. In addition, the capacity of magnetic tweezers to manipulate a particle in 3D is limited. The use of more than two magnetic poles increases maneuverability, as shown in recent reports [13], but 3D control of the particle can remain a problem for many experiments.

Particles that do not show a magnetic response cannot be manipulated directly with magnetic fields. Cells and organelles, for example, cannot be dragged with magnetic tweezers. While this is a limitation for *in vitro* experiments, this fact is being exploited in the manipulation of magnetic nanoparticles inside a cell with high specificity [13].

Magnetic and optical tweezers, together with atomic force microscopy (AFM) (Chap. 8), cover a range of forces from 10^{-2} to 10^{5} pN for single molecule manipulation. These techniques thus provide modern means to look into biochemical reactions in new ways. In addition to the research scope and experimental conditions, spatial and temporal resolution issues should be considered to aid in making the most appropriate choices [14].

9.4 Operation

Forces on biological specimens of a few nanometers are very small, and it is thus difficult to generate a stable trap in liquid media. Thermal fluctuations and instrument noise make trapping efficiency very low and residence times very short. It is also difficult to control specimen orientation and to assure that a single specimen is in the trap. The methodology for manipulating nucleic acids, viruses and proteins therefore involves the use of beads as mechanical handles. We will now explain general strategies for performing experiments in the optical tweezers.

9.4.1 Methodology

Most experiments in the optical tweezers require that the molecular system under study be tethered to at least two specific sites. For this purpose, a second bead must be introduced in the experiments, either by suction by a micropipette (shown in Fig. 9.4c, bottom), or using a second optical trap. As stated above, these two beads are mechanical handles to biochemically link, for example, DNA molecules by two ends, or more complex systems such as a virus particle and the viral DNA to be packaged (see below). To exert force, one of the beads must move relative to the other (Fig. 9.4c). In optical designs in which the second bead is held by suction, the micropipette is moved by a computer-controlled stage. In designs in which this micropipette is replaced by another optical trap, the second trap can be displaced by steering one laser beam relative to the other with external mirrors and lenses (Fig. 9.5).

In DNA stretching experiments, the molecule is often modified at one end to insert biotin labels and at the other to insert digoxigenin (Dig) labels. The molecule is then tethered between two beads, one coated with streptavidin to provide biotin-streptavidin linkage and the other with anti-digoxigenin (α-Dig) to provide a Dig-α-Dig linkage. The experiment is followed by videomicroscopy, observing the position of the two beads. DNA extension is measured by recording the relative positions of the two beads in real time. In setups in which a micropipette is used, a low-power laser can be introduced to track the position of the second bead; light scattered from the bead can be detected in another photodetector. Alternatively, if the micropipette rigidly attached to the microchamber is moved relative to the

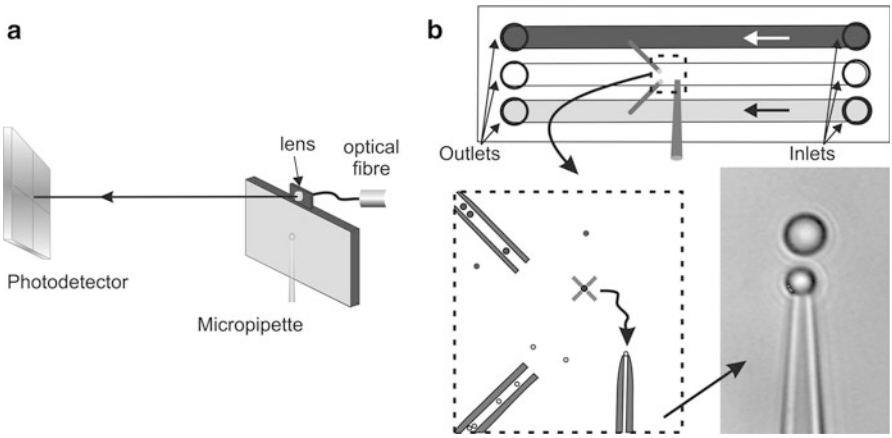


Fig. 9.6 Optical tweezers methodology. (**a**) Use of a light-lever device to track the position of the chamber with the rigidly attached micropipette. (**b**) *Top*: a three-channel microfluidics chamber is filled with buffer. Two suspensions of different types of microspheres are then flowed into the top and bottom channels through inlet holes; the fluid exits the channels through outlet holes. *Bottom*, *left*: microspheres from the top or bottom channel are optically trapped at the exit of the dispenser tubes and deposited on the micropipette in the middle of the central channel. *Bottom*, *right*: micrograph as seen by videomicroscopy of a ~3-μm diameter bead in the optical trap and a ~2-μm diameter bead held by the micropipette

optical trap (Fig. 9.6a), a so-called light-lever device can be used [6]. This device comprises a single-mode laser and a lens attached to the pipette/chamber frame as in Fig. 9.6a. The lens collimates light from the laser and directs it onto a photodetector. A similar light-lever configuration is used in AFM to detect the position of the cantilever in reflection (see Chap. 8).

The procedure for a DNA stretching experiment in the fluidics chamber is shown in Fig. 9.6b. The optical trap is first used to grab a single streptavidin-coated microsphere from one of the channels and to place it on the micropipette. An α-Dig-coated bead with attached DNA fibers is then trapped from the other channel and brought toward the bead on the micropipette until a stable biotin-streptavidin bond is formed.

9.4.2 Molecular Construct

The design of the molecular construct can be a bottleneck in single-molecule optical tweezers experiments. Good design consists of biochemically engineering the specimen for attachment to mechanical handles. Force is a vector magnitude and therefore marks a precise coordinate in the experiments, unlike other parameters such as temperature or pH. It is important to tether the specimen correctly from the domains that will produce mechanical work or will extend.

To study the elasticity of a polymer, for example DNA, the molecular construct is very simple and consists of tethering the molecule from opposite ends, as explained above. If one wants to unzip double-stranded (ds) DNA to study mechanically induced melting, the molecule must be tethered on opposite strands of the same end. When proteins are involved, structural data are needed to determine the sites expected to unfold. In experiments with molecular motors, structural data and functional models are desirable, to establish the domains involved in motion and the direction the molecule is expected to translate or rotate.

The design of the molecular construct can be seen as the analogue of the crystallization procedure in X-ray diffraction (Chap. 4): proteins need different approaches for their structural analysis with this technique and established strategies help, but do not constitute a standard for new experiments. Much of the success in single-molecule experiments relies on correct construct design. Figure 9.7 shows a molecular construct devised to study bacteriophage packaging by *Bacillus subtilis* phage Φ29. Phage capsids with prepackaged 6.6-μm-long dsDNA were bound directly to a bead on a micropipette, tethering the free DNA end to an optically trapped bead [17]. We will analyze later in this chapter how this construct is used to study the Φ29 packaging mechanism.

9.4.3 Operation Modes

There are four basic operation methods: *force-extension*, *constant-force*, *passive* and *force-jump* modes. The first is used to stretch polymers for analysis of their elastic behavior; Fig. 9.7 shows the force-extension curve of a DNA molecule in this mode. Polymer extension is observed as the force increases continuously; for this reason, this mode is also known as force-ramp.

The constant-force or force-feedback mode is used to analyze the activity of a protein machine as a function of a constant load. In this mode, the position of the optical trap relative to the second attachment point is adjusted to maintain constant tension on the bead. The protein kinesin has been tested using this strategy, as has packaging activity of Φ29 (Fig. 9.7) (find particular examples in [1]). Application of distinct constant forces allows observation of varying responses of the specimen.

The passive mode consists of fixing the position of the micropipette and/or traps and letting the specimen work. In this mode, the distance between the beads is not necessarily constant, as the position of the trapped bead undergoes small displacements relative to its resting position, resulting from the activity of the tethered specimen. This mode can be used to the study of DNA polymerization, kinesin transport and viral packaging, and allows measurement of maximum forces developed by protein motors, also known as stall forces.

In force-jump mode, the applied force is changed rapidly in discrete force steps rather than continuously (characteristic of the ramp-mode). This mode is useful for studying kinetics between several molecular states that are stable at different forces.

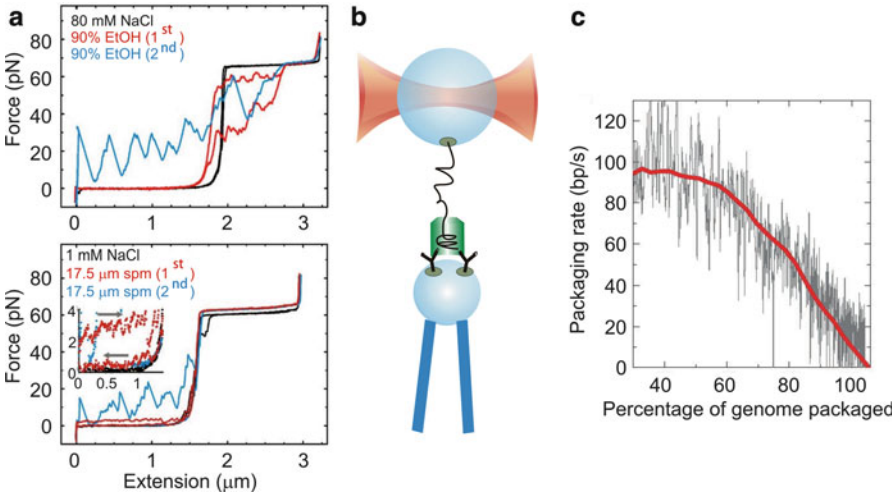


Fig. 9.7 Some applications of optical tweezers. (**a**) DNA stretching experiments. In each graph, a single molecule of dsDNA is first stretched and relaxed in physiological buffer (*black curve*), followed by two successive stretch/relax cycles (*red* and *blue curves*) in a buffer-90 % ethanol mixture (*top*) or in buffer with a low concentration of spermine (*bottom*) (Reproduced from [15]. With permission). (**b**) Molecular construct to study Φ29 packaging activity. The bead in the optical trap is streptavidin-coated for attachment to the biotin-labeled free end of the DNA. The virus is attached to a protein G-coated bead held by a micropipette. The DNA-capsid complex is captured in the tweezers in a stalled, partially prepackaged preparation before the start of the assay. (Current strategies are not restricted to using prepackaged molecular constructs [16]) (**c**) Packaging rate *versus* the amount of DNA packaged, relative to the original 19.3-kbp Φ29 genome. *Grey line*, trace for a single complex; *red line*, average of eight such measurements (Reproduced from [17] with permission)

For instance, protein folding or RNA-hairpin open/close conformations can be monitored through changes in the molecular extension with this operation mode [18].

9.5 Understanding Viruses: Some Major Contributions of Optical Tweezers

In contrast to the techniques studied in previous chapters (Chaps. 3, 4, 5, 6, 7 and 8), viruses have been little studied with optical tweezers so far. The reason is the novelty of the single-molecule approaches that involve mechanical manipulation of biological specimens. There is currently much interest not only in characterizing viral mechanochemical functions in real time, but also in using this machinery for nanotechnological applications. In this section, we present some major and illustrative contributions provided by optical tweezers in virology.

9.5.1 *Elasticity of Nucleic Acids and Relevance for Packaging in Viruses*

With the advent of optical tweezers in molecular biology, DNA and RNA in their double- or single-stranded configurations were promptly studied [19]. The first assays in the 1990s elucidated the elasticity of these information carriers. These studies are important for understanding the energetic barriers and the function of proteins and molecular complexes that act on nucleic acids in processes such as replication, transcription, unwinding or packaging. From the point of view of structural and physical virology, conformational dynamics of nucleic acids is important, for example, for understanding nucleic acid packaging and storage inside viral capsids, from passive encapsulation into preassembled viral capsids to the active packaging process in bacteriophages (see also Chap. 12).

A molecular construct for mechanical analysis of DNA is shown in Fig. 9.4c. Figure 9.7a shows the results of a standard experiment in which individual dsDNA molecules are stretched and then relaxed in different conditions. For these experiments, DNA was obtained from a linear fragment of plasmid piJ702 from *Streptomyces*, whose makeup consists of 70 % guanine-cytosine (GC) base pairs (bp). Each molecule was first stretched and relaxed in physiological conditions (aqueous buffer with 80 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5) and subsequently stretched and relaxed twice in a buffer-90 % ethanol mixture (Fig. 9.7a, top) or in the same buffer with 1 mM NaCl and 17.5 μM spermine (Fig. 9.7a, bottom). The ability of optical tweezers to deal with different conditions in the same assay is demonstrated here. Ethanol was used to mimic the behavior of DNA at low humidity, an environment similar to that inside viral capsids; spermine, a polycation involved in genome condensation, was used to study DNA organization in free form, not constrained by, for example, encapsulation in viral proheads.

The force-extension curves in Fig. 9.7a in physiological buffer (black trace) show the characteristic elasticity regimes of DNA: the entropic regime, in which the molecule behaves as an entropic spring, and the intrinsic regime, in which it is stretched beyond its contour length. At 65 pN, the molecule overstretches as a result of the almost total unwinding it experiences at high stress [19]. Force-extension curves of DNA in the ethanol mixture or with spermine show either a plateau at entropic forces of ~3 pN or a stick-release pattern. These effects indicate the unraveling of collapsed structures, such as coiled DNA or toroidal supercoiling, in the form of a smooth decondensation process or a violent unfolding of the polymer (see [15] and references therein).

The mechanical properties of DNA, as characterized by its persistence length and stretch modulus of the molecules, have been also studied as a function of the GC composition or of external conditions such as pH, ionic strength, temperature and humidity level (find advanced material in [1, 20–22]). Optical tweezers have also been used in more varied configurations to study DNA unzipping (the mechanical separation of the two strands), DNA twisting (with optical and magnetic tweezers), and the opening of RNA hairpins [18, 19]. The influence of several

DNA-binding proteins on the mechanical properties of the complex has also been analyzed [23].

In conclusion, from the point of view of viral nucleic acid storage and packaging in viral capsids, these and other single-molecule manipulation studies have contributed to the understanding of the elastic, electrostatic and entropic penalties of confining and organizing genetic material at high pressure in a small, closed space. These studies have also shown that nucleic acid structure is dependent on mechanical tension; as indicated, a few piconewtons are sufficient to straighten a dsDNA molecule; some tens of piconewtons are sufficient to promote base-pair unstacking and twist changes in physiological conditions, both of which are mechanical denaturation effects that lead to geometric changes in the DNA double helix. Forces in these ranges are developed by the bacteriophage portal machinery, which will be reviewed in the next section. A dynamic description of viral packaging in the presence of mechanical stress must therefore include the tension-induced conformational alterations of the viral nucleic acid that the portal machinery must process.

9.5.2 Packaging Activity of Bacteriophages

Bacteriophages are the most abundant organisms on earth (see Chap. 17). Nonetheless, very few virologists have analyzed their machinery at the single particle level [24–26]. The pioneering work of Smith and Bustamante's team in 2001 reported the direct activity of this impressive, powerful machinery [17]. These results have complemented structural and biochemical studies (see Chap. 12) and also opened entirely new perspectives on the working of viral nanomachines. An artistic view of a typical bacteriophage manipulation assay in the optical tweezers is represented in Fig. 9.8.

Measurement of the portal motor of bacteriophage Φ29 in the passive mode showed that its packaging machinery is highly processive (*i.e.* that can undergo many catalytic cycles without dissociation) and one of the strongest known in nature. Forces against an external load (like that exerted by a micropipette, Fig. 9.7b) or by another optical trap (Fig. 9.9a) can be above 60 pN. These forces are near those that distort the dsDNA in the overstretching transition (see force-extension curves in Fig. 9.7a). Constant-force mode studies indicated that packaging rates, which can reach 100 base pairs per second (100 bp/s), are punctuated by pauses and slips, and that these rates are dependent on the percentage of total genome packaged (see Fig. 9.7c).

Since that time, the Smith and Bustamante labs have contributed to understanding the mechanochemical behavior of this machinery. The former extended the use of optical tweezers to other phages in an assay similar to that in Fig. 9.7b. Studies of T4 [29] and λ [30] phages showed that high force generation is a common property of viral packaging motors. Among other results, this group found that T4 works at a strikingly high translocation rate, on average 700 bp/s, with maximum rates of

Fig. 9.8 A bacteriophage packaging assay in the optical tweezers. This artist's conception depicts a bacteriophage Φ29 packaging its genome against a force applied from a laser-trapped bead. The viral head is attached to the upper bead by chemical bonds represented by chains. The *in vitro* experiment is performed in aqueous buffer. A fibreless Φ29 virion map (cross section) at 7.8-Å resolution, obtained from EMdatabank.org (ID: emd-1,420 [27]), has been modified for illustrative purposes

2,000 bp/s [29]. In λ phage, procapsid expansion (see Chap. 13) was detected at 30 % of total packaging [30]. The Smith group also analyzed ionic effect on packaging rates and on portal motor function [31], as well as pressure within the capsids at different stages of the packaging process. They developed a new strategy to analyze the packaging process at near-zero force, which also involves initiation of the process without using a prepackaged configuration [16].

The dependence of the internal force on capsid filling is also a relevant analysis for its implication in the capsid pressure. The Smith team reported internal forces exceeding 100 pN, which are also indicative of the resistance of the capsid shells; these forces can be characterized in indentation experiments using AFM (see Chaps. 8 and 18). One of these studies showed that the capsid can withstand indentation forces up to ~1,000 pN [32], a result compatible with internal forces of 100 pN after complete capsid filling [16]. In addition, Smith and co-workers also analyzed the effect of mutations in the phage λ motor on its packaging dynamics [33, 34].

The Bustamante group determined the kinetic parameters of the Φ29 packaging motor and their dependence on external load. These researchers also showed that

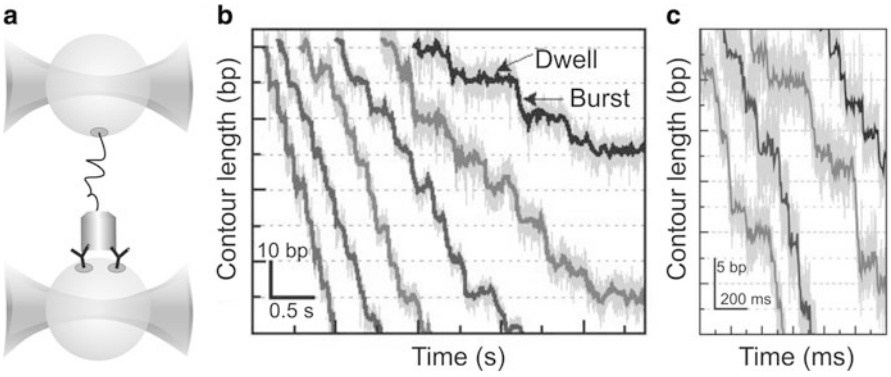


Fig. 9.9 High resolution measurements of packaging activity. Detecting a structure in motion on increasingly smaller scales involves a simultaneous improvement on spatial and temporal resolutions. (**a**) A dual-trap optical tweezers in which two separate traps are built from the same laser beam allow better sensitivity, as detailed elsewhere [7, 8]. (**b**) Step size in Φ29. The graph shows representative packaging traces collected under low external load ~8 pN, and different [ATP] (from *left* to *right*): 250, 100, 50, 25, 10 and 5 μM, all boxcar-filtered and decimated to 50 Hz. Data at 1.25 kHz are plotted in *light grey*. Packaging is observed to occur in large 10-bp "bursts" separated by flat "dwells". (**c**) Fine structure of the 10-bp bursts: each step consists of four 2.5-bp substeps. The graph shows representative packaging traces collected with external loads of ~40 pN and 250 μM [ATP], all boxcar-filtered and decimated to 100 Hz. Data in *light grey* are plotted at 1.25 kHz (Reproduced from [28] with permission)

the motor subunits act in a coordinated, successive fashion with high processivity [35]. Their study suggests that the packaging rate decrease shown in Fig. 9.7c at high capsid fillings results from build-up of internal pressure due to DNA confinement, which exerts an opposing force and slows the motor. Recent studies from this lab characterized the intersubunit coordination of the homomeric ring ATPase, which powers the DNA packaging process in Φ29. ATPase characterizations of this kind are important, because this machinery is also involved in other varied tasks in the cell, from chromosome segregation to protein degradation [28]. For these studies, the Bustamante lab used high-resolution optical tweezers (see Sect. 9.6). They found that packaging occurs in increments of 10 bp, in turn composed of four 2.5-bp steps, as shown in Fig. 9.9. An important conclusion is that a non-integer number of base-pair steps can take place in protein mechanics, in contrast to earlier models that assumed an integer number. In another assay using several modified DNAs, these researchers dissected phosphate interactions in the DNA-motor engagement and showed that promiscuous, nonspecific contacts with the substrates are common in translocase-substrate interactions [36].

Promiscuity by viral DNA packaging machinery was also recently studied by Rao and colleagues [37], who reported similar packaging activity rates for proheads and previously filled heads of T4 bacteriophage. Full heads and those emptied of most of their packaged DNA can reassemble the packaging machinery and reuse it. This group also found that several exogenous DNA molecules can be used as a substrate to fill the T4 capsid in several initiation steps [37].

The use of magnetic tweezers to provide a tethering force over a single DNA-phage complex was used in conjunction with single molecule fluorescence. In this experiment, Bustamante and colleagues tested the possibility of connector rotation in Φ29 [38]. Using an impressive variety of instrumentation and techniques, they showed that absolute rotations do not take place; only transient rotation of the connector relative to the capsid shell, followed by return to the original position, is likely to be involved in the packaging process.

The mechanical properties and structural transitions of nucleic acids in the presence of force are useful information to devise and test DNA confinement models, as prompted in the previous subsection. Several theoretical models have been proposed to understand DNA organization inside the capsids (see [25] and references therein). These models together with the above experimental analysis on single viruses are needed to understand several aspects of the relationship between a virus and its substrate; for example, the connection between internal pressure and DNA ejection forces.

Finally, single-molecule manipulation studies have provided new insights into the theoretical modeling of bacteriophage packaging mechanisms, which would be limited if based only on information from structural studies, solution biochemistry and genetics [39, 40]. For example, the possibility that DNA rotates during packaging, a process compatible with that in which the connector does not rotate can be directly tested. (A translocation mechanism driven by a (non-rotating) compression/extension mechanism is not supported by experimental evidence). The connection between DNA structure and the viral packaging mechanism at high force can also be enlightened. Although the portion of DNA engaged in the ATPases is closely in B-form due to structural constraints in the connector, the DNA fragment outside the connector and capsid should strongly deviate from a double helix at forces above 65 pN. It is thus tempting to speculate that ATPases might have trouble in doing good contacts with the DNA polymer above 65 pN thus decreasing the packaging efficiency. This fact might be one of the reasons why average maximum packaging forces are around 50–60 pN in Φ29. In order to test this scenario, constant forces above 65 pN should be directly applied against the motor.

9.5.3 Manipulation of Individual Viruses and Other Configurations

Direct trapping of individual virions, without using mechanical handles, is possible and has been shown using the tobacco mosaic virus in one of the pioneering manipulation experiments by Ashkin. The insights into the field of virology provided by these studies were limited, however, and this method has not found broad applications so far [41]. However, large viruses of hundreds of nanometers are amenable to individual examination in an optical trap, and this setup can be used to test direct responses to external physical stimuli (electric fields, local heat, or a

combination of these with chemical changes such as those of pH and ionic strength), as shown for other biological structures [9]. Motion against fluid force can be used to characterize capsid friction and hydrodynamic size. The physical characterization of virus particles may be important not only to understand virus function, but also for nanotechnological applications of virus particles (see Chap. 22).

Virus-cell attachments have also been characterized by force-measuring optical tweezers. In these experiments, a laser-trapped bead coated by influenza virions was approached to a host cell to measure the interactions between an individual virus and the cell receptors. Rupture forces in the range of 10–25 pN and sequential unbinding events were detected and confirmed by similar AFM assays. Molecular dynamics simulations (see Chap. 19) were complementarily used to explain unbinding pathways and interaction dynamics [42]. In another configuration, cell-cell transmission of HIV-1 virus through a synapse was demonstrated by manipulation of single cells with optical tweezers [43].

9.5.4 A Note on Viral Nucleic Acid Replication and Transcription

The replicative protein machinery of viruses, which involve a variety of force-developing motors such as polymerases and helicases, is the object of a growing number of mechanistic studies at the single-molecule level using optical tweezers. These studies exceed the scope of this book; however, interested readers will find illustrative contributions in the literature [44–46].

9.6 Perspectives

There is little debate today as to the importance of viewing processes that take place inside cells as a product of nanoscale machinery. Proteins, nucleic acids and biomolecular complexes are not only chemical entities, they are known to perform mechanical work, an activity inherently related to their function. Manipulation techniques have become key instruments for identifying the mechanochemical activity of this biological machinery. These techniques are thus used not only as confirmatory tests for chemical and structural models, but also address novel mechanisms that cannot otherwise be determined [3].

Optical tweezers are one of the most important tools in this endeavor, and have furnished continuous advances since Ashkin's pioneering work in 1970 [47]. In the previous section Sect. (9.5) and in Chemla and Smith's chapter [25], some perspectives for the study of viruses themselves are discussed. Here we focus on optical advances and hybrid instruments.

9.6.1 Improvements in Resolution

One of the most important challenges in the field of biomolecular manipulation is to increase resolution in measurements. In bacteriophages, motors that translocate or move nucleic acids are inextricably coupled to their geometric structure. They use these motors to encapsulate nucleic acids to near crystalline densities inside a nanoscopic container. DNA is a double-helical sequence with a distance between base pairs of 0.34 nm in B-form. Resolving discrete steps in DNA packaging can involve spatial resolution of single base pairs, which is limited by drift and noise. Noise can be that intrinsic to the experiments, caused by fundamental Brownian fluctuations, and extrinsic, caused by instrumental and environmental disturbances (power fluctuations of the lasers, electronic noise, acoustic and mechanical room vibrations, etc.). A drift in micropipette position or changes in optical components (refractive index variations or thermal expansion/contraction of lenses and mirrors) due to temperature gradients, for example, can alter the calibration. Noise and drift can thus lead to false signals or reduce the accuracy of measurements.

Temporal resolution is another issue that is continually under development. Unlike structural studies in which the specimen is static, dynamic studies involve detection of conformational changes in structures in motion on their most fundamental temporal scale. Resolution is actually a combined function of space and time in optical tweezers and thus, improvements in resolution entail a simultaneous advance on spatial and temporal detection. Improved sensitivity can uncover the discrete stepping structure of a coordinated machine, as shown in Fig. 9.9 for the Φ29 packaging mechanism [28]; in this example, the 10-bp steps of each translocation burst (Fig. 9.9b) was shown to be composed of four 2.5-bp substeps (Fig. 9.9c) by a high-resolution instrument [8].

Recent advances have pushed spatial and temporal resolution to the Angstrom and millisecond scales, respectively, with measurement stability within the second range. The use of two traps from the same laser beam [8] has specifically allowed direct observation of intersubunit coordination and of step size in the packaging motor of Φ29 [28]. Future progress is open to new ideas for specific treatment of each experimental noise source; for example, the use of smaller beads as mechanical handles decreases the effects of Brownian noise, although they are technically more difficult to manipulate [7].

9.6.2 Adding Translation Coordinates and Rotation

Measurement in more degrees of freedom constitutes another challenge. The result of several coordinated molecular motors is not only linear displacement, but also a general three-dimensional motion that might involve several translations and rotations. Multiple optical traps permit measurement of motion in more than one linear coordinate, as well as control over more than one molecule [48]. To date,

molecular specimens have been biochemically linked by two specific sites to beads in one or two traps, but the use of other configurations with more than two traps was shown, for example, to allow simultaneous manipulation of several molecules in multiple directions [49].

Another development in manipulation that could help to understand viral packaging processes is the combination of torque measurements in the optical tweezers. Torque is easily produced in magnetic tweezers (see Sect. 9.3.7) and has been generated in optical tweezers using a rotating micropipette [19]; however, torque measurements are technically more difficult. Recent advances point toward use of the intrinsic properties of light [50]. Light can exchange angular momentum -spin or orbital angular momentum, or both- and can thus exert rotation on particles. Conversely, angular momentum changes in the light scattered from a particle can be analyzed as a method to measure torque [51].

The helical structure of DNA suggests that its binding machinery follows its helical pitch to generate rotation. As discussed above, DNA may move by ratcheting and rotation in the Φ29 phage portal motor in its translocation into the viral capsid, a mechanism that could be examined by a combination of fluorescence microscopy and optical tweezers (see below). Torque-measuring optical tweezers could be then applied to characterize torsion-induced mechanochemical packaging pathways by over- and underwinding the DNA double helix.

9.6.3 Combination with Fluorescence Imaging

Imaging of molecules while they are being manipulated is a long-standing demand. There is currently no method with which to perform structural studies with atomic resolution of a molecule in motion or under mechanical stress. The conditions involved in methods such as X-ray diffraction or NMR are incompatible with those used in liquid chambers. It is nonetheless possible to introduce methods developed for optical microscopes in the tweezers setup.

Combining fluorescence with optical tweezers not only provides control over the experiment, but also allows observation of new conformations induced by force or torque. Single-molecule fluorescence and confocal imaging have already been incorporated successfully to optical tweezers, and their applicability is shown by a number of studies (see [7, 52] and references therein). A general strategy to integrate fluorescence detection in optical tweezers was outlined briefly in Fig. 9.5 [53–55]. Fluorescence can be obtained using fluorophores or with quantum dots, which add stability in signal emission. The use of a pair of dyes allows detection of nanoscale distance changes of a protein or of a complex of coordinated proteins using Förster resonant energy transfer (FRET) photophysical interactions [56].

Fluorescence has been exploited to understand DNA structure under tension as well as the portal motor dynamics of bacteriophages [38, 52]. These studies suggest a fruitful combination of methods to visualize bacteriophage motor dynamics under

the control of force and torque. DNA could be fluorescently labeled to follow its twisting dynamics relative to the connector in the packaging experiment.

9.6.4 Conclusions

We have introduced a technique with which the mechanical action of viruses as well as the elastic properties and condensation dynamics of nucleic acids can be analyzed. Optical tweezers are a laser-based tool with which to trap and manipulate individual particles with piconewton forces; they are used in molecular biology to measure pairs of associated magnitudes such as force-extension and torque-rotation as a function of time. The resulting dynamic information from these experiments is yielding not only demonstrative experiments of previously conjectured mechanistic models, but also unparalleled characterization of the viral machinery. Since the experiments are performed in a liquid chamber, optical tweezers can be adapted to a large variety of physiological conditions. They can also be merged with optical microscopies, allowing information from manipulation experiments and real-time imaging to be integrated in the same experiment.

The dynamic information provided by optical tweezers and its combination with structural and biochemical analysis provide rigorous ways to model viral function and stability. These models in turn permit more precise thermodynamic characterization of the viral machinery, and form a basis for the development of nanotechnological applications.

Acknowledgements It is a pleasure to acknowledge J.R. Moffitt and J.L. Carrascosa for technical insights into different aspects of the chapter, R. Bocanegra, L. Quintana for careful reading of the manuscript, C. Mark and S. Hormeño for editorial and illustration assistance, respectively, and M. de la Guía for graphic design of Fig. 9.8. This work was funded by the Spanish Ministry of Science and Innovation under the "Ramon y Cajal" program (Grant No. RYC-2007-01765).

References and Further Reading

1. Hormeno S, Arias-Gonzalez JR (2006) Exploring mechanochemical processes in the cell with optical tweezers. Biol Cell 98:679–695
2. Svoboda K, Block SM (1994) Biological applications of optical forces. Annu Rev Biophys Biomol Struct 23:247–285
3. Bustamante C (2008) In singulo biochemistry: when less is more. Annu Rev Biochem 77:45–50
4. Oster G, Wang H (2003) How protein motors convert chemical energy into mechanical work. In: Schliwa M (ed) Molecular motors. Wiley-VCH, Weinheim, pp 207–227
5. Ketterle W (1999) Experimental studies of Bose-Einstein condensation. Phys Today 52:30–35
6. Smith SB, Cui Y, Bustamante C (2003) Optical-trap force transducer that operates by direct measurement of light momentum. Methods Enzymol 361:134–162
7. Moffitt JR, Chemla YR, Smith SB, Bustamante C (2008) Recent advances in optical tweezers. Annu Rev Biochem 77:205–228

8. Moffitt JR, Chemla YR, Izhaky D, Bustamante C (2006) Differential detection of dual traps improves the spatial resolution of optical tweezers. Proc Natl Acad Sci U S A 103:9006–9011
9. Hormeno S, Ibarra B, Chichon FJ, Habermann K, Lange BM, Valpuesta JM, Carrascosa JL, Arias-Gonzalez JR (2009) Single centrosome manipulation reveals its electric charge and associated dynamic structure. Biophys J 97:1022–1030
10. Jackson JD (1999) Classical electrodynamics. Wiley Hoboken, USA
11. Arias-Gonzalez JR, Nieto-Vesperinas M (2003) Optical forces on small particles: attractive and repulsive nature and Plasmon-resonance conditions. J Opt Soc Am A 20:1201–1209
12. Tanase M, Biais N, Sheetz M (2007) Magnetic tweezers in cell biology. In: Yu-Li W, Dennis ED (eds) Methods in cell biology, vol 83. Academic Press, pp 473–493, vol 83
13. Neuman KC, Nagy A (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods 5:491–505
14. Bustamante C, Macosko JC, Wuite GJ (2000) Grabbing the cat by the tail: manipulating molecules one by one. Nat Rev Mol Cell Biol 1:130–136
15. Hormeno S, Moreno-Herrero F, Ibarra B, Carrascosa JL, Valpuesta JM, Arias-Gonzalez JR (2011) Condensation prevails over B-a transition in the structure of DNA at low humidity. Biophys J 100:2006–2015
16. Rickgauer JP, Fuller DN, Grimes S, Jardine PJ, Anderson DL, Smith DE (2008) Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage phi29. Biophys J 94:159–167
17. Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. Nature 413:748–752
18. Li PT, Collin D, Smith SB, Bustamante C, Tinoco I Jr (2006) Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. Biophys J 90:250–260
19. Bustamante C, Bryant Z, Smith SB (2003) Ten years of tension: single-molecule DNA mechanics. Nature 421:423–427
20. Williams MC, Rouzina I, Bloomfield VA (2002) Thermodynamics of DNA interactions from single molecule stretching experiments. Acc Chem Res 35:159–166
21. Mao H, Arias-Gonzalez JR, Smith SB, Tinoco I Jr, Bustamante C (2005) Temperature control methods in a laser tweezers system. Biophys J 89:1308–1316
22. Hormeno S, Ibarra B, Valpuesta JM, Carrascosa JL, Arias-Gonzalez JR (2012) Mechanical stability of low-humidity single DNA molecules. Biopolymers 97:199–208
23. Chaurasiya KR, Paramanathan T, McCauley MJ, Williams MC (2010) Biophysical characterization of DNA binding from single molecule force measurements. Phys Life Rev 7:299–341
24. Casjens SR (2011) The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol 9:647–657
25. Chemla YR, Smith DE (2012) Single-molecule studies of viral DNA packaging. Adv Exp Med Biol 726:549–584
26. Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. Adv Exp Med Biol 726:489–509
27. Tang J, Olson N, Jardine PJ, Grimes S, Anderson DL, Baker TS (2008) DNA poised for release in bacteriophage phi29. Structure 16:935–943
28. Moffitt JR, Chemla YR, Aathavan K, Grimes S, Jardine PJ, Anderson DL, Bustamante C (2009) Intersubunit coordination in a homomeric ring ATPase. Nature 457:446–450
29. Fuller DN, Raymer DM, Kottadiel VI, Rao VB, Smith DE (2007) Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. Proc Natl Acad Sci U S A 104:16868–16873
30. Fuller DN, Raymer DM, Rickgauer JP, Robertson RM, Catalano CE, Anderson DL, Grimes S, Smith DE (2007) Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. J Mol Biol 373:1113–1122

31. Fuller DN, Rickgauer JP, Jardine PJ, Grimes S, Anderson DL, Smith DE (2007) Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi 29. Proc Natl Acad Sci U S A 104:11245–11250
32. Ivanovska IL, de Pablo PJ, Ibarra B, Sgalari G, MacKintosh FC, Carrascosa JL, Schmidt CF, Wuite GJ (2004) Bacteriophage capsids: tough nanoshells with complex elastic properties. Proc Natl Acad Sci U S A 101:7600–7605
33. Tsay JM, Sippy J, DelToro D, Andrews BT, Draper B, Rao V, Catalano CE, Feiss M, Smith DE (2010) Mutations altering a structurally conserved loop-helix-loop region of a viral packaging motor change DNA translocation velocity and processivity. J Biol Chem 285:24282–24289
34. Tsay JM, Sippy J, Feiss M, Smith DE (2009) The Q motif of a viral packaging motor governs its force generation and communicates ATP recognition to DNA interaction. Proc Natl Acad Sci U S A 106:14355–14360
35. Chemla YR, Aathavan K, Michaelis J, Grimes S, Jardine PJ, Anderson DL, Bustamante C (2005) Mechanism of force generation of a viral DNA packaging motor. Cell 122:683–692
36. Aathavan K, Politzer AT, Kaplan A, Moffitt JR, Chemla YR, Grimes S, Jardine PJ, Anderson DL, Bustamante C (2009) Substrate interactions and promiscuity in a viral DNA packaging motor. Nature 461:669–673
37. Zhang Z, Kottadiel VI, Vafabakhsh R, Dai L, Chemla YR, Ha T, Rao VB (2011) A promiscuous DNA packaging machine from bacteriophage T4. PLoS Biol 9:e1000592
38. Hugel T, Michaelis J, Hetherington CL, Jardine PJ, Grimes S, Walter JM, Falk W, Anderson DL, Bustamante C (2007) Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids. PLoS Biol 5:e59
39. Purohit PK, Kondev J, Phillips R (2003) Mechanics of DNA packaging in viruses. Proc Natl Acad Sci U S A 100:3173–3178
40. Yu J, Moffitt J, Hetherington CL, Bustamante C, Oster G (2010) Mechanochemistry of a viral DNA packaging motor. J Mol Biol 400:186–203
41. Ashkin A, Dziedzic JM (1987) Optical trapping and manipulation of viruses and bacteria. Science 235:1517–1520
42. Sieben C, Kappel C, Zhu R, Wozniak A, Rankl C, Hinterdorfer P, Grubmuller H, Herrmann A (2012) Influenza virus binds its host cell using multiple dynamic interactions. Proc Natl Acad Sci U S A 109:13626–13631
43. McNerney GP, Hubner W, Chen BK, Huser T (2010) Manipulating CD4+ T cells by optical tweezers for the initiation of cell-cell transfer of HIV-1. J Biophoton 3:216–223
44. Herbert KM, Greenleaf WJ, Block SM (2008) Single-molecule studies of RNA polymerase: motoring along. Annu Rev Biochem 77:149–176
45. Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. Annu Rev Biophys 37:317–336
46. van Oijen AM, Loparo JJ (2010) Single-molecule studies of the replisome. Annu Rev Biophys 39:429–448
47. Ashkin A (1970) Acceleration and trapping of particles by radiation pressure. Phys Rev Lett 24:156–159
48. Grier DG (2003) A revolution in optical manipulation. Nature 424:810–816
49. Dame RT, Noom MC, Wuite GJ (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature 444:387–390
50. Gutierrez-Medina B, Andreasson JO, Greenleaf WJ, Laporta A, Block SM (2010) An optical apparatus for rotation and trapping. Methods Enzymol 475:377–404
51. Parkin S, Knoner G, Singer W, Nieminen TA, Heckenberg NR, Rubinsztein-Dunlop H (2007) Optical torque on microscopic objects. Methods Cell Biol 82:525–561
52. van Mameren J, Peterman EJ, Wuite GJ (2008) See me, feel me: methods to concurrently visualize and manipulate single DNA molecules and associated proteins. Nucleic Acids Res 36:4381–4389

53. Lang MJ, Fordyce PM, Engh AM, Neuman KC, Block SM (2004) Simultaneous, coincident optical trapping and single-molecule fluorescence. Nat Methods 1:133–139
54. Richardson AC, Reihani N, Oddershede LB (2006) In: Dholakia K, Spalding GC (eds) Combining confocal microscopy with precise force-scope optical tweezers, vol 6326. SPIE, San Diego, pp 632628–632637
55. Vossen DLJ, van der Horst A, Dogterom M, van Blaaderen A (2004) Optical tweezers and confocal microscopy for simultaneous three-dimensional manipulation and imaging in concentrated colloidal dispersions. Rev Sci Instrum 75:2960–2970
56. Toprak E, Selvin PR (2007) New fluorescent tools for watching nanometer-scale conformational changes of single molecules. Annu Rev Biophys Biomol Struct 36:349–369

Further Reading

Berns MW, Greulich KO (2007) Laser manipulation of cells and tissues. Methods Cell Biol 82
Schliwa M (ed) (2003) Molecular motors. Wiley-VCH Weinheim (Germany)
Sheetz MP (1998) Laser tweezers in cell biology. Methods Cell Biol 55

Also especially recommended for further reading are references [1, 7, 13, 24, 25] *listed above*.

Part III
Structural Foundations of Virus Properties and Functions

Chapter 10
Assembly of Simple Icosahedral Viruses

José M. Almendral

Abstract Icosahedral viruses exhibit elegant pathways of capsid assembly and maturation regulated by symmetry principles. Assembly is a dynamic process driven by consecutive and genetically programmed morphogenetic interactions between protein subunits. The non-symmetric capsid subunits are gathered by hydrophobic contacts and non-covalent interactions in assembly intermediates, which serve as blocks to build a symmetric capsid. In some cases, non-symmetric interactions among intermediates are involved in assembly, highlighting the remarkable capacity of capsid proteins to fold into demanding conformations compatible with a closed protein shell. In this chapter, the morphogenesis of structurally simple icosahedral viruses, including representative members of the parvoviruses, picornaviruses or polyomaviruses as paradigms, is described in some detail. Icosahedral virus assembly may occur in different subcellular compartments and involve a panoplia of cellular and viral factors, chaperones, and protein modifications that, in general, are still poorly characterized. Mechanisms of viral genome encapsidation may imply direct interactions between the genome and the assembly intermediates, or active packaging into a preformed empty capsid. High stability of intermediates and proteolytic cleavages during viral maturation usually contribute to the overall irreversible character of the assembly process. These and other simple icosahedral viruses were pioneer models to understand basic principles of virus assembly, continue to be leading subjects of morphogenetic analyses, and have inspired ongoing studies on the assembly of larger viruses and cellular and synthetic macromolecular complexes.

J.M. Almendral (✉)
Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM) and Department of Molecular Biology of the Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain
e-mail: jmalmendral@cbm.uam.es

M.G. Mateu (ed.), *Structure and Physics of Viruses; An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_10,

Keywords Icosahedral capsid • Triangulation number • Assembly intermediate • Protein contacts • Nuclear translocation • Protein folding • Nucleation • Packaging • Encapsidation • Virus factory • Maturation cleavage

Abbreviations

AAP	assembly-activating protein
AAV	adeno-associated virus
CBB	capsid building block
CP	capsid protein
CPV	canine parvovirus
DBD	DNA-binding domains
H1-PV	parvovirus H1
hr-t	host range-transforming
FPV	feline parvovirus
MEV	mink enteritis virus
MVM	minute virus of mice
NLM	nuclear localization motif
NLS	nuclear localization sequence
NPC	nuclear pore complex
PPV	porcine parvovirus
SV40	simian virus 40
VLP	virus-like particle
VP	viral protein
5x	five-fold axis
3x	three-fold axis
2x	two-fold axis.

10.1 Introduction

All viral entities have a capsid, which in structurally simple viruses is built up from one or a few types of protein subunits. At late stages of the intracellular phase of their life cycle (see Chap. 1) viruses perform capsid assembly, a process by which the structural capsid protein (CP) subunits are joined by maximal hydrophobic contacts and/or non-covalent interactions (and occassionally covalent bonds) to construct the viral particle. This process is essential for viruses to mature (become infectious) and release progeny, and in the simple icosahedral viruses proceeds by strict principles of genetic economy and symmetry (see Chap. 2). This chapter reviews the main stages in the assembly and genome encapsidation of small or medium sized viruses with a relatively simple icosahedral architecture, exemplified by three distinct virus models: (i) Adeno-associated virus (AAV) and minute virus

of mice (MVM), as respective representative members of the *Dependovirus* and *Parvovirus* genera of the *Parvoviridae*, a family of single-stranded (ss) DNA viruses with a T = 1 capsid (25 nm in diameter) assembling in the nucleus; (ii) Poliovirus, genus *Enterovirus* of the *Picornaviridae*, RNA(+) viruses with a pseudoT = 3 capsid (30 nm in diameter) assembling in the cytoplasm; and (iii) Polyomavirus and Simian Virus (SV40), members of the *Polyomavirus* genus of the *Polyomaviridae*, double-stranded (ds) DNA viruses with an all-pentamers T = 7d capsid (45 nm in diameter) assembling in the nucleus. In these so-called "simple" icosahedral viruses, assembly occurs through an orchestrated pattern of interactions of the capsid subunits to form complexes or assembly intermediates, whose composition and conformation usually change along the process. The structural dynamics undergone by the assembly intermediates must also fulfill another important function, which is to traffic within the infected cell towards the compartment where the genome is being replicated. It is the accumulation of assembly intermediates at a specific compartment what triggers genome encapsidation and maturation, allowing the virus to finally propagate in nature.

10.2 Icosahedral Capsids: Symmetry and Genetic Regulation

10.2.1 Structural Principles in Icosahedral Capsid Assembly

In simple icosahedral viruses, the capsid is formed by many copies of one or a few protein subunits that assemble by making multiple contacts to build a hollow shell of proper size and symmetry. The regular icosahedron is formed by a defined number of copies of a capsid building block (CBB), which can be built up by a single CP subunit, by several identical CPs, or by non-identical CPs. The CBBs are related by two-fold (2×), three-fold (3×), and five-fold (5×) symmetry axes, and CPs within the CBBs establish regular interactions with their neighbors depending on their position in relation to these icosahedral axes. The Caspar and Klug (1962) theory explained how some multiples of 60 identical subunits could be arranged with similar (quasi-equivalent) interactions, according to the rule $T = h^2 + hk + k^2$, where h and k are integers and T is called the triangulation number (see Chap. 2 for a detailed explanation of capsid icosahedral symmetry and quasi-equivalence).

In the members of the *Parvoviridae*, the capsid is a T = 1 perfect icosahedron, formed by a total of 60 CPs (termed viral proteins, VPs, in this and some other virus families), which include two to three variant types (VP1, VP2 and VP3) with identical amino acid sequence and fold except for short streches of sequences at the C- or N-termini, which are intrinsically disordered and are not observed in the X-ray structures of the virus particles (Fig. 10.1a). The topology of the capsid surface differs among the parvoviruses due to the characteristic prominence of peptide loops and spikes at the 3x symmetry axes, and the depth and contour of the depression surrounding the 5x axes [1, 2, 7–10]. In Poliovirus and related picornaviruses the protein shell is built up of 60 copies of a fundamental subunit (the protomer) composed of three different proteins (termed VP1, VP2 and VP3)

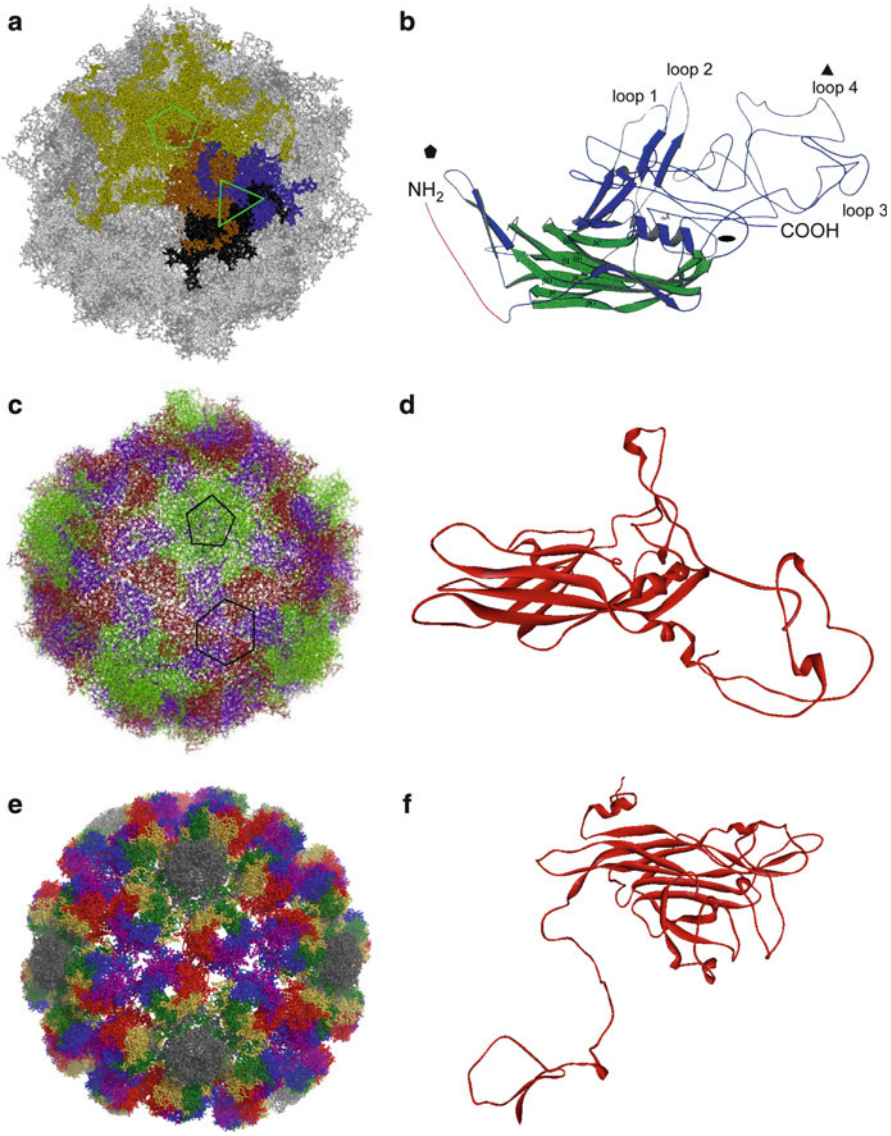


Fig. 10.1 Atomic structure of icosahedral virus particles and capsid subunits whose assembly is described in this chapter. (**a**) Structure of Parvovirus MVM (p and i strains; [1, 2]). Different *colours* distinguish the subunits surrounding the 5x axes (*pentagon*) from those interdigitated at the 3x axes (*triangle*). (**b**) Folding of the VP1 and VP2 subunits in the MVM structure. Note the prominent loops projecting away from the capsid surface. (**c**) Structure of Poliovirus, a picornavirus [3, 4]. The VP1 subunits around the 5x axes (*pentagon*), and the alternating VP2 and VP3 subunits around the 3x axes (*hexagon*) are respectively shown in *green, red* and *violet*. (**d**) Folding of the VP1 subunit in the capsid of Poliovirus. The β-strands (*arrows*) forms two antiparallel sheets juxtaposed in a wedgelike structure. (**e**) Structure of the SV40 virus showing the organization of the VP1 pentamers [5]. The 5x-coordinated pentamers are represented in gray and the VP1

which are not related in amino acid sequence but have a similar fold, and a small extended polypeptide (VP4). VP4 is located at the inner surface of the protein shell and remains covalently linked to the N-terminus of VP2 until the final stages of virus assembly and maturation. The interactions between the VP1 subunits around the 5x axes are not equivalent to those engaging the alternating VP2 and VP3 subunits in the 3x axes, accounting for the features of the surface of the capsid (Fig. 10.1c), including the conspicuous protrusion of the VP1 subunits at the 5x axes [3, 4]. In polyomaviruses such as SV40, the icosahedral capsid is formed by 360 identical CP subunits of a single protein (termed VP1), whose arrangement does not follow the quasi-equivalence rules of Caspar and Klug, as the basic structural elements (capsomers) are 72 pentamers displayed in a T = 7d surface lattice ([5]; Fig. 10.1e). Each pentamer contains in addition one copy of either of two other proteins (termed VP2 and VP3), which share most of their amino acid sequences. The capsomers located around each 5x axis are referred to as the pentavalent pentamers, and those capsomers arranged around each of the 3x axis are referred to as the hexavalent pentamers.

Although the parvovirus, poliovirus and polyomavirus capsid proteins are not related in amino acid sequences, the core of all these proteins are folded in their capsids as a β-sheet structure termed β-barrel jelly roll (or Swiss-roll β-barrel, or an eight-stranded antiparallel β-barrel), a wedge-shaped structure comprising two antiparallel β-sheets. The topology of all these capsid subunits is, thus, similar (Fig. 10.1b, d and f). The major structural differences between them are in the loops that connect the strands, which are particularly prominent in parvoviruses (Fig. 10.1b), and in the N- and C- terminal segments that extend from the central β-barrel domains. The remarkable similarity of these β-barrel jelly rolls among viral proteins that do not share primary amino acid sequences, and belong to different families of viruses with unrelated biological properties, suggests that it may represent one of the few structural solutions allowing proteins to be packaged in icosahedral capsids, or be a testimony of a common ancestral evolutionary history (see also Chaps. 2 and 7).

10.2.2 Synthesis of the Capsid Subunits: Setting the Assembly Scenario

For successful assembly, the synthesis of the capsid proteins must be tightly regulated during the virus life cycle, in order to satisfy at least three key requirements: (i) quantity, viruses must induce the accumulation of high amounts of capsid

Fig. 10.1 (continued) subunits of the pentamers in hexameric arrays are represented in different *colors*. (**f**) folding of the VP1 subunit as observed in the SV40 capsid. The β-barrel (jellyroll) is radial to the capsid surface. The figure was prepared using the Pymol program (http://www.pymol.org/) and the VIPER resources [6], using the atomic coordinates of MVM (1MVM, 1Z1C), Poliovirus (2PLV), and SV40 (1SVA), deposited in the Protein Data Bank (PDB)

proteins in the host cells, in order to compete with the vast amount and diversity of preexisting cellular proteins; (ii) timing, capsid proteins synthesis must reach maximum levels by the time the viral genome is being replicated, so it can be efficiently packaged; and (iii) stoichiometry, capsid subunits must be synthesized at the proper ratio to ensure assembly of an infectious viral particle. Genetic regulation of viral gene expression is therefore crucial for a successful assembly process. Mechanisms controlling viral gene expression involve multiple networks which are out of the scope of this book, so only a brief outline of those involved in assembly are mentioned below.

In Poliovirus (see [11] for a review of its life cycle), the genomic RNA is released in the cytoplasm by the incoming virion, and translated as a single open reading frame to produce a very large polyprotein, which undergoes cotranslational *cis*-cleavage by the viral protease 2A. This cleavage releases from the N-terminus the precursor polyprotein myristoyl-P1, which contains the CP sequences. In this and related picornaviruses, CPs are synthesized to high levels by the combination of the translational competence of the genomic RNA with an effective host protein shut-off induced by viral proteases, whereas the proper protein stoichiometry results from the cleavage of a common precursor.

In parvoviruses, polyomaviruses and other DNA viruses, the stoichiometry of capsid proteins is regulated mainly at the level of splicing of their messenger RNAs. In the parvovirus MVM for instance, site-directed mutagenesis at the minor splicing sites, or independent cDNA cloning, allows to obtain genomic clones expressing either VP protein [12]. VP2 alone can form a capsid which can encapsidate the viral genome, but VP1 is necessary for the infectivity of the particles due to specific domains residing at its N-terminal segment [13, 14]; and (Fig. 10.3c). A 1:5 ratio of VP1:VP2 is found for the soluble synthesized proteins, as well as in the assembled capsid, and preserving this ratio is critically important for an ordered assembly avoiding protein aggregates [16]. In polyomaviruses VP1, VP2, and VP3 levels in the infected cells are regulated by alternative splicing from a common trascript, occurring soon after viral DNA replicative intermediates accumulate. In these small DNA viruses protein shut-off is not a major mechanisn to counteract host protein synthesis, but the nuclear accumulation of protein products prior assembly (see below) facilitates their interactions at certain nuclearly confined enviroments.

10.3 Capsid Building Blocks and Assembly Intermediates

The capacity of the viral structural proteins to form capsids ultimately result from their folding and self-assembly properties, which are conferred by the encoded amino acid sequences. Virus capsids are assembled from CBBs which may be CP monomers or, in many cases, CP oligomers. These stable oligomeric CBBs may be considered as the first stable intermediates of the capsid assembly process. The use of *in vitro* assembly systems in combination with diverse theoretical approaches has allowed the investigation of fundamental principles of virus capsid self-assembly

starting from CBBs. The theory of capsid self-assembly is outside the aims of this chapter, which is dedicated to the description of assembly processes in different viruses from *in vitro* and *in vivo* experimental evidences supported by structural and functional studies. The reader is referred to Chap. 1 for a brief overview, and to Chap. 19 for a detailed description on the thermodynamic and kinetic aspects of the assembly of simple virus capsids. These studies are generally consistent with experimental observations on the assembly of very simple virus capsids, frequently carried out in controlled *in vitro* experiments. The models support assembly as a nucleated cooperative process in which CP or CBB concentration is critical, with a lag phase reflecting the time required to build up an assembly line of intermediate structures. The intermediates are expected to be at very low concentration, but this steady state of intermediates is required for efficient assembly in any stepwise reaction.

10.3.1 Structure of CBBs and Assembly Intermediates

The stepwise assembly pathway has been well characterized in Poliovirus (and other picornaviruses including human rhinovirus and foot-and-mouth disease virus), as discrete intermediates are stable enough to be isolated. A common strategy that many viruses adopt to build blocks for capsid assembly is to initiate assembly while the structural units are linked into a polyprotein precursor. In Poliovirus, the first intermediate of the assembly pathway is an inmature structural unit (unprocessed protomer) formed by a folded polyprotein (termed P1). P1 contains three structural domains. These domains are split apart from each other upon cleavage at specific sites in the linker sequences by the viral $3CD^{pro}$ protease. The result is a processed protomer (5S protomer) which sediments as a 5S particle and is formed by one copy each of VP0, VP3, and VP1, which correspond to the cleaved P1 domains (Fig. 10.2a). It is unclear when the β-barrel of these proteins core fold, but unprocessed P1 of picornaviruses is recognised by panels of virus-neutralising antibodies elicited against discontinuous epitopes of mature virions, strongly suggesting that the domains in unprocessed P1 are already folded like the mature CPs in the assembled capsid. The structures of the isolated unprocessed or processed P1 protomers have not been solved for any picornavirus yet; however, in the virus capsid structure VP1, VP2 and VP3 form an intricate network of intermolecular interactions among the surfaces of their β-barrel domains, which must greatly stabilize the protomer and contribute to the early stages of the assembly. VP4 remains covalently linked to VP2 in the VP0 protein until virus assembly is completed and maturation occurs (see below).

The 5S precursor is followed in the assembly line by the 14S pentamer (Fig. 10.4a), which is formed by oligomerization of five 5S protomers, and is stabilized by extensive protein-protein interactions and by others mediated by myristate chains incorporated in the five VP0 N-termini. These multiple interactions determine a molecular interlocking of the five protomers in the pentamer, conferring high stability and directionality to the whole assembly pathway.

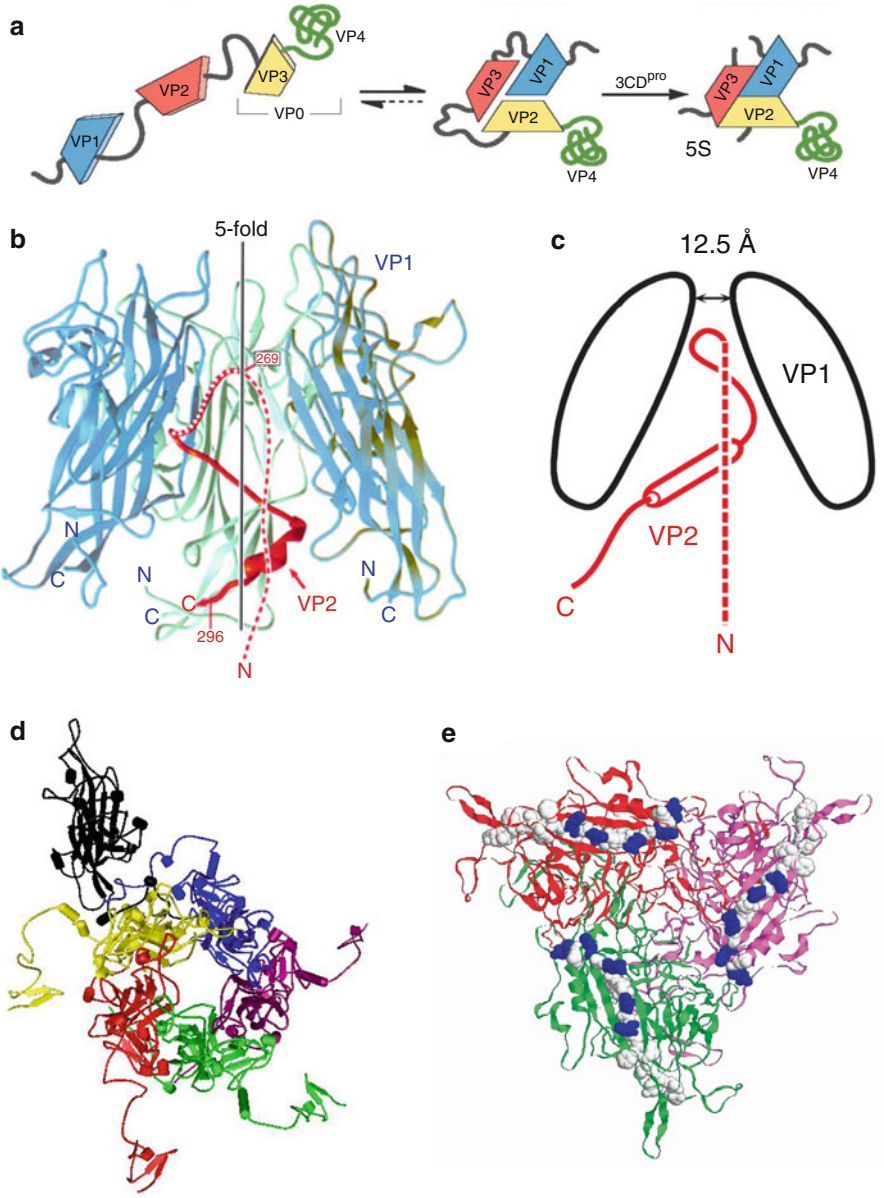


Fig. 10.2 Structural models of assembly intermediates in icosahedral viruses. (**a**) Poliovirus P1 polyprotein precursor showing the four proteins forming the heterometric structural unit (protomer). Cleavage by $3CD^{pro}$ protease yields the 5S assembly intermediate. (Adapted from Flint et al. (2009) Principles of Virology, ASM Press, with permission). (**b**) Structure of the VP1-VP2 assembly intermediate of polyomavirus. VP2 is shown in red and the three VP1 monomers that form contacts with VP2 in *green* (*middle*) and *blue* (*left* and *right*). (**c**) Schematic representation of the VP1/VP2 interaction in this complex. VP2 (*red*) enters in the VP1 pentamer (*black*) from the base, continuing to the *upper* part of the conical depression. It then loops back to interact

In polyomaviruses as well as in parvoviruses, the CP subunits interact in the cytoplasm before nuclear capsid formation. The composition of the assembly intermediates is characteristic for each virus system and plays a role of paramount importance in the correct timing and spatial ordering during assembly. A relatively stable cytoplasmic assembly intermediate seems to be a common need for these viruses completing capsid assembly and maturing within the nucleus. In Polyomavirus and SV40, stable pentamers of VP1 are formed in the cytoplasm. Each VP1 pentamer binds either one VP2 or one VP3 protein that becomes allocated in the axial cavity of the pentamer (Fig. 10.2b), and the VP1-VP2/3 complex is stabilized by strong hydrophobic interactions [17]. The contacts between subunits significantly alter the configuration of the VP1 pentamer, as demonstrated by changes in epitope accesibility and by direct structural insights obtained from the crystal structure (Fig. 10.2b, c). In SV40, transient disulphide bridges are established intramolecularly, and subsequently intermolecularly, as the monomers assemble into pentamers, which facilitate the folding and interdigitation of structural elements [19]. The noncovalent interactions and covalent bonds collectively conform a stable cytoplasmic CP pentameric complex (Fig. 10.2d), which is the major assembly intermediate in these viruses.

In parvoviruses, taking the murine MVM as a reference model, trimers of VP subunits assemble in the cytoplasm. The VP1 (82 kDa) and VP2 (63 kDa) proteins synthesized at a VP1:VP2 1:5 M ratio assemble into two types of trimers (Fig. 10.2e), which are produced in stoichiometric amounts. The larger trimer (200 kDa) is a heterotrimer formed by one VP1 and two VP2 subunits, whereas the smaller (180 kDa) is a VP2-only homotrimer [16]. In the formation of the cytoplasmic trimer, the VP2 protein may act as a scaffolding factor assisting VP1 to acquire a proper folding, as deletion mutants of the VP1-specific region undergo extensive ubiquitination degradative reaction that can be significantly prevented by co-expression of VP2 [14]. The assembly of VP cytoplasmic trimers for these viruses was structurally supported first by the higher stability of the trimer (measured by the buried surface area on oligomer formation) as compared to putative dimers or pentamers of CP subunits, due to the multiple contacts in the intertwined loops of the subunits around the 3x symmetry axes (Fig. 10.1a). Furthermore, crosslinking experiments and mutations disrupting the intertrimer interfaces in the MVM capsid allowed the isolation of stable trimers (see below).

Fig. 10.2 (continued) specifically with the inner face of VP1, forming a hairpin-like structure. (**b**, **c**: Reprinted from [17], with permission). (**d**) Structure of the VP1 pentamer in the SV40 capsid. (**e**) Structure of the VP trimer in the MVM capsid. Residues involved in NLM function (see *below*) are highlighted. (**d**, **e**: prepared using the Pymol program (http://www.pymol.org/) and the 1SVA and 1MVM respective atomic coordinates deposited in the PDB)

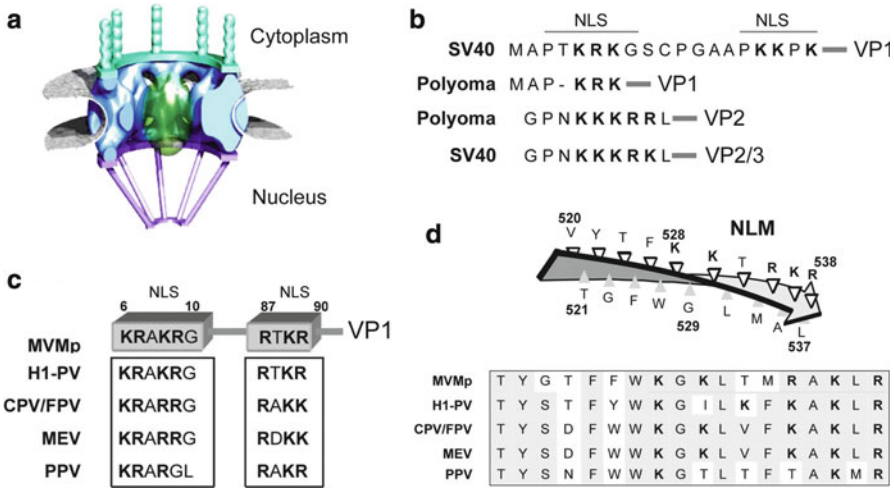


Fig. 10.3 Nuclear localization sequences in polyomavirus and parvovirus capsid proteins. (**a**) Basic architecture of the nuclear pore complex (NPC). The central ring is illustrated with the filament facing the cytoplasm and the basket protruding into the nucleus. (**b**) Examples of identified NLS in the VP proteins of Polyomavirus and SV40. (**c**) Two conserved domains with NLS activity identified in the VP1 N-terminal sequence of several parvoviruses: MVM, parvovirus H1 (H1-PV), CPV, feline parvovirus (FPV), mink enteritis virus (MEV), porcine parvovirus (PPV). (**d**) Configuration of the NLM in the capsid subunits of MVM and conservation of the NLM sequence in related parvoviruses. Basic residues contributing to nuclear targeting of assembly intermediates are shown in *bold*. (Adapted from [15] with permission)

10.3.2 Intracellular Traffic of CBBs and Assembly Intermediates

For capsid formation, CBBs or other assembly intermediates must accumulate in the subcellular compartment where the viral genome replicates, and at the right time. In some viral systems, the capsid subunits must traffic from the site of synthesis in the cytoplasm to the assembly compartment. This traffic is directed by protein signals in CBBs or intermediates that are accessible to the transport machineries of the cell. The nature of the signals and the configuration of the intermediate exposing them, which will determine the transport route accessed by the intermediate and its fate in the cell, are thus key elements in the viral assembly process.

In polyomaviruses and parvoviruses, the viral genome is replicated in the nucleus, and the CPs of these viruses are therefore karyophilic polypeptides that traverse the nuclear membrane. Translocation of proteins into the nucleus imposes two restrictions, firstly by the need of nuclear localization sequences (NLS) to access the cellular transport machinery, and secondly by the size of the complex to be transported, which cannot exceed 25–30 nm, the functional diameter of the aperture in the nuclear pore complex (NPC) [20], a supramolecular structure embedded in the nuclear envelope (Fig. 10.3a). The exchange of macromolecules

across the NPC is mostly mediated by proteins of the importin β (karyopherin β) superfamily, which comprises importins and exportins. The NLS are, in the so-called conventional configuration [21], single or bipartite stretches of basic amino acids that access, generally upon direct binding to a protein adaptor, the importin α/β transport pathway [22]. However, many karyophilic proteins harbor non-classical NLS that may bind karyopherin β1 directly, or access the alternative karyopherin β2/transportin import pathway. A current active area of research is the identification of the transport routes accessed by the viral proteins in specific cell hosts. Functional conventional NLS are found in the major VP1 as well as in the minor VP2/3 subunits of Polyomavirus [23, 24] and SV40 [25] (Fig. 10.3b). It is remarkable the sequence conservation among the structural proteins of these viruses at the NLS domains, compared to the high sequence divergence in the rest of these proteins. In polyomaviruses the VP1-VP2/3 cytoplasmic complex is translocated through the NPC by the functional cooperation of the multiple NLS displayed by the VP subunits [26].

In MVM, used as a representative molecular model of the *Parvoviridae*, the two types of VP trimers are major CBBs, or stable assembly intermediates, translocating across the nuclear membrane. The protein subunits within the trimer cooperate for nuclear transport, as both VP1 and VP2 proteins genetically depleted of functional nuclear transport sequences can be co-transported into the nucleus by expressed intact subunits [14]. Indeed VP1 and VP2 carry independent NLS and efficiently target the nucleus of transfected cells when singly expressed [12]. VP1 harbours two conventional NLS at its N-terminal specific region (Fig. 10.3c), which function independently. These NLS are required for nuclear translocation of the expressed VP1 subunits in MVM [14] and in canine parvovirus, (CPV) [27], and also for the MVM virion to initiate infection [14], suggesting that they are exposed out of the virus shell during the cell entry process. Similar separate regions with basic amino acids essential for assembly and infectivity were identified in the capsid proteins of AAV [28]. In addition, both VP1 and VP2 (the major capsid-forming polypeptide) contain, in their common folded sequence, a structured domain with nuclear targeting capacity (named NLM) [15]. In the assembled capsid the NLM is localized in the amphipatic ß-strand I at the inner capsid surface (Fig. 10.3d). The NLM is the only functional nuclear targeting sequence identified in VP2. The structured NLM, highly conserved in many members of the *Parvoviridae*, shows, when displayed in the capsid structure (Fig. 10.2e), all its charged basic amino acids placed within the side of the strand facing the solvent at the interior surface of the capsid, and the hydrophobic amino acids oriented towards the protein core. This structuration of the NLM probably occurs only upon folding and trimerization of the VPs, which may occur concomitantly. Folding and oligomerization of the capsid subunits probably occur in the cytoplasm, leading to trimeric assembly intermediates acquiring nuclear transport competence. The structuration of the two types of trimers translocating across the NPC may exert a quality control role in the virus morphogenetic flow, as misfolded subunits (not exposing the NLM) or oligomers with aberrant structures and protein composition would not access the nucleus [15, 16].

10.4 Forming the Capsid

10.4.1 Contacts and Structural Changes in CBBs and Assembly Intermediates

The CBBs or initial stable intermediates of capsid assembly may be found in an assembly-incompetent state, incapable of establishing productive interactions between them to accommodate the final configuration of the capsid [29]. Thus, the assembling oligomers must frequently undergo conformational changes during the late stages of the assembly pathway. In Poliovirus and other picornaviruses, the similar β-barrel folding of the VP1, VP2 and VP3 proteins that facilitate their interactions to form the 60 structural subunits (5S protomers) and the 14S pentameric CBBs (Fig. 10.4a), also favours the subsequent assembly of these latter intermediates into a complete viral particle [30]. For this, the extensive interactions among the β-barrels of adjacent proteins help to form a rather rigid protein shell with a dense network of intersubunit interactions. Inside the capsid, a network of additional protein contacts stabilizes the mature particle. These contacts are largerly contributed by the long (40–80 residues) N-terminal arms of the three VP subunits, which have a similar path inside the capsid in different picornaviruses, although their primary amino acid sequence is drastically different. These interactions are most extensive at the 5x axes, where the N-termini of five VP3 molecules are arranged in a tubelike parrallel β-sheet. The conformation of these arms is ordered only after capsid assembly, so it is difficult to evaluate the contribution of the interactions involving these arms in the final particle stability based on structural data only.

In members of the *Parvoviridae*, capsid formation in the natural infection starts in the nucleus. Indeed the nucleolus was identified as the subcellular compartment where AAV assembly inititates, colocalizing with the non-structural replicative viral Rep proteins; subsequently, capsid accumulation spreads across the entire nucleus [31]. In MVM, as the assembly intermediates (mainly trimers) accumulate in the nucleus, the nucleation reaction is triggered and multiple non-covalent interactions are established between amino acids localized at the edges of the binding trimers. A few evolutionary conserved residues involved in presumably strong intertrimer contacts were found to be necessary for capsid assembly (Fig. 10.4c). These residues buried a large hydrophobic surface upon trimer association, or formed buried intertrimer hydrogen bonds or salt bridges [18]. Assembly intermediates other than the CBBs are difficult to isolate in parvoviruses, presumably because of the high efficiency of the assembly reaction in which intermediates are highly transient and accumulate at very low levels, in agreement with theoretical studies on the assembly of simple virus capsids (see Chap. 19). A genetic approach to trap unstable assembly intermediates based on the introduction of disruptive mutations at some of the residues that in the capsid are involved in contacts between MVM trimers yielded high amounts of trimers, without evidence of any other larger intermediate [16]. These trimers could be isolated by

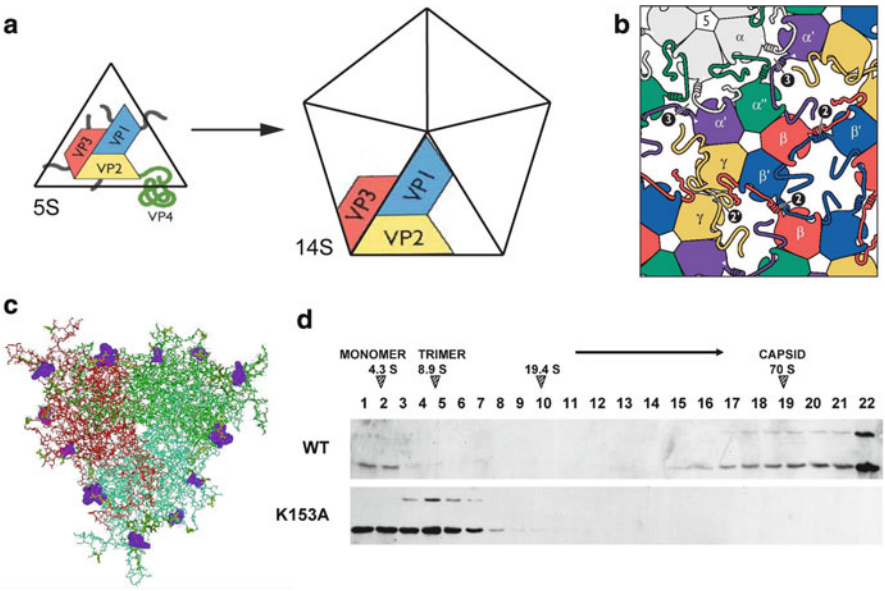


Fig. 10.4 Interactions among assembly intermediates to form the capsid. (**a**) Assembly of the 5S protomer into a 14S pentamer in Poliovirus. The protomer is not identical to the icosahedral asymmetric unit in the capsid. (**b**) Contacts between capsomers in SV40. Four pentamers assembled in the SV40 capsid are shown. Each subunit of the pentamers projects an arm that makes different contacts with the jellyroll of the subunit in another pentamer. (Adapted from [5], with permission). (**c**) Capsid formation in Parvovirus: a VP trimer is represented; some residues involved in major inter-trimer contacts of MVM capsid and required for capsid assembly are shown as spacefill models and coloured *violet*. (Adapted from [18], with permission). (**d**) Sedimentation analysis in sucrose gradients of the K153A mutant of the MVM capsid proteins. This mutant lacks a side chain that is important for establishing critical contacts between MVM trimers during capsid assembly. The sedimentation position of complete capsids, trimers, and VP monomers are indicated. (Adapted from [16], with permission)

sedimentation in sucrose gradients (Fig. 10.4d), and were competent in a nuclear transport assay performed in permeabilized cells [32], indicating that their configuration properly expose the nuclear transport sequences, and thus mutations precluded just the final capsid formation stage.

The acquisition of competence for assembly by MVM trimers involves conformational changes that have been indirectly detected. An MVM-induced monoclonal antibody recognizing a discontinuous epitope located at each 3x axis of the capsid (which corresponds to the center of each trimer) failed to react with non-assembled trimers accumulated in the cell nucleus or isolated *in vitro* [16, 32]. Trimers must therefore change their conformation during nuclear capsid assembly, and the process may be triggered by some external factor and/or phosphorylation of the subunits (see below). These induced conformational rearrangements would reorganize some residues located at the vertex of the thee-fold axes, creating a new epitope.

In the viral particle of Polyomavirus and SV40, all the capsomers are pentamers, even though the capsid shell is built by subunits arranged in an icosahedral T = 7 lattice. Therefore the architecture of this virus does not fullfil the arrangement of pentameric and hexameric capsomers predicted by the quasiequivalent theory for capsids with T >1 values ([33]; see Sect. 10.2.1). However the SV40 capsid is stable enough, and this is accomplished in part by virtue of unique intercapsomers bonds established during assembly by the C-terminal domain of the VP1 protein (Fig. 10.4b). This domain acts as a connecting arm which, under six different configurations, makes contacts between the pentavalent pentamers (around each 5× axis) and the hexavalent pentamers (around each 3× axis). The different types of contacts in which the arms are involved act as multiple clamps holding the subunits together, and thus ensuring the stability of the capsid shell [5].

The VP1 protein of Polyomavirus can self-assemble into pentamers in heterologous cell expression systems [34], and the purified pentamers can associate to form capsid-like assemblies (VLPs) that become stabilized at low ionic strength by calcium. Interestingly, these VLPs assemble without the minor virion protein components, VP2 and VP3, suggesting that the non-equivalently related subunits of the penta- and hexavalent capsomeres may spontaneously switch their bonding specificity during assembly. VP1-only capsids however do not represent the physiological assembly pathway observed in the natural virus infections, in which the assembly of the VP1 pentamers is driven upon regular interactions involving VP2 and VP3 with the viral genome DNA (see below), and empty capsids are thought to be minor abortive assembly by-products.

10.4.2 Compartments, Factors, and Protein Modifications Influencing Capsid Formation

Icosahedral capsid assembly is influenced by environmental conditions in the cell and also by multiple molecular factors (in addition to the VP protein themselves) that may be encoded by the cells or by the virus genomes. These factors may act at specific subcellular compartments, and their contribution may critically determine in many cases the efficiency of the assembly process *in vivo*. Some of these factors have been identified in several viral systems, although multiple experimental evidences suggest that many others remain to be identified. For instance in Poliovirus, the rate of assembly of the structural proteins *in vitro* is reduced by at least two orders of magnitude compared to the rate observed in infected cells, and the empty capsids formed showed altered conformation, unless the reaction is primed by 14S pentamers isolated from infected cells. This experiment implies that proper folding, interactions, and/or modifications of the proteins forming the pentameric 14S intermediate, are essential for assembly to proceed successfully. A candidate for this function is the cellular Hsp70 chaperone, which associates with the P1 polyprotein precursor during its folding to form the 5S structural unit.

The cellular modulations of virus assembly may involve complex signalling pathways. For parvoviruses, capsid assembly during natural infections occurs with high efficiency in the nuclear compartment of infected host cells. However virus-like particles (VLPs) devoid of nucleic acid may be formed in the cytoplasm of heterologous expression systems (*e.g.* recombinant baculovirus expressing VP2 of MVM in insect cells) at low efficiency [35, 36]. This distinct assembly efficiency may be accounted, at least in part, by post-translational modifications of the capsid subunits. In MVM-infected cells, VPs and native capsid become extensively post-translationally modified by phosphorylation [37], whereas VLPs purified from heterologous insect cells were not phosphorylated [32]. In spite of the absence of modifications, these VLPs showed a 3-D structure identical to the native capsid and virus [2, 35], indicating that phosphorylation is not important for the icosahedral T = 1 ordering. The 2D-tryptic phosphopeptides analysis of native MVM capsid subunits resulted in a complex pattern of phosphoserine and phosphothreonine residues which was different for VP1 compared to VP2 [37]. In the host cell systems studied, the phosphorylation of the VP subunits of MVM was mainly catalyzed by the cytoplasmic activity of the Raf-1 kinase of the MAPK signalling pathway [32], and this modification was crucial for the acquisition of nuclear transport competence by the trimers. Phosphorylation by cellular kinases may be a general strategy to connect capsid assembly with host cell physiology, ensuring a spatially and timely regulated process maximizing virus yield.

As in parvoviruses, in Polyomavirus the efficiency of nuclear assembly is regulated by phosphorylation of the capsid subunits. Polyomavirus is highly tumorigenic in mouse, but host range transforming (hr-t) mutants of this virus defective in tumour induction are blocked in virion assembly when infecting non-permissive cells [38], although viral DNA and capsid proteins are synthesized to wild type levels. In purified Polyomavirus particles the several VP1 species identified by 2-D electrophoresis are generated by acetylation and phosphorylation of threonine and serine residues of the initial translation product. The hr-t mutants failed to assemble the complete (240S) viral particle, correlating with the lack of acidic forms of VP1 [39]. These acidic forms resulted from phosphorylation of threonine residues, and at least one of the phosphotreonines was shown to be essential for the encapsidation of the viral minichromosome [45]. In SV40, phosphate groups are added in natural infections on serine and threonine residues flanking the NLS of the VP proteins. Phosphorylation was shown to act on the activity of these NLS indirectly regulating the nuclear accumulation of VPs and virus assembly.

An example of a factor encoded by the viral genome favouring assembly was identified in studies with the parvovirus AAV. Capsid assembly in AAV begins in the nucleolus, and spreads throughout the nucleus at later stages of infection. In addition to the three capsid proteins, VP1, VP2 and VP3, the *cap* gene also encodes, by using an alternative open reading frame, the assembly-activating protein (AAP) that is essential for capsid assembly [31]. The AAP factor targets newly synthesized capsid proteins to the nucleolus, and becomes stabilized upon co-expression of the capsid protein VP3 of the same virus serotype, or from an AAV serotype of the same assembly group. The assembly-promoting activity of AAP is mediated by

interaction between two hydrophobic domains in the N-terminal region of the molecule with the C-terminus of the VP proteins, which forms the capsid protein interface at the 2× symmetry axes [40]. AAP seems to act as a scaffolding factor able to change the conformation of non-assembled VP molecules.

10.5 Genome Encapsidation and Virus Maturation

10.5.1 Poliovirus Cytoplasmic Maturation

Pathways of genome packaging adopted by some icosahedral viruses have been, and continue to be a matter of debate, as it is quite difficult in many cases to distinguish between "concerted assembly", in which the capsid is formed as a result of the ordered association of the protein subunits with the genome, from the "sequential assembly" that occurs when the genome is encapsidated into a preformed capsid, which may require an active stage (see Chap. 12 for detailed descriptions of mechanisms of encapsidation referred mostly to structurally complex viruses). Figure 10.5a illustrates this dilema in Poliovirus. In this system, most evidences support a "concerted" assembly of the 14S pentamer condensing around the RNA genome. Alternatively, an empty procapsid containing 60 copies of the VP0-VP3-VP1 structural unit could be transiently required at low concentrations for packaging. The apparent equilibrium between pentamers and an empty capsid of low stability makes it difficult to demonstrate whether the RNA is encapsidated by association with the pentamers or becomes inserted into the capsid [11].

Whichever encapsidation pathway is dominant, the maturing Poliovirus virion must undergo a number of modifications prior to becoming infectious. Maturation involves lipid modification and proteolytic cleavage at a specific site, making assembly an irreversible process. Along the final stages of assembly, a molecule of the fatty acid myristate is added post-translationally to the N-terminus of each VP4 subunit. This lipid mediates the interaction of the β-sheet formed by VP3 N-termini with a second β-sheet structure containing strands contributed by both VP4 and VP1 molecules. This feature of the capsid does not form until final stages of the maturation, when proteolytic processing liberates VP2 and VP4 from their precursor VP0, and this reaction is associated with a significant increase in the stability of the viral particle, priming it for entry into a new host cell.

10.5.2 Polymorphic Nuclear Maturation of Polyomaviruses

To form the SV40 and Polyomavirus virions, the VP1-pentamers become associated with single copies of a minor capsid protein (either VP2 or VP3) and, once imported into the nucleus, interact with the replicating DNA. Although the

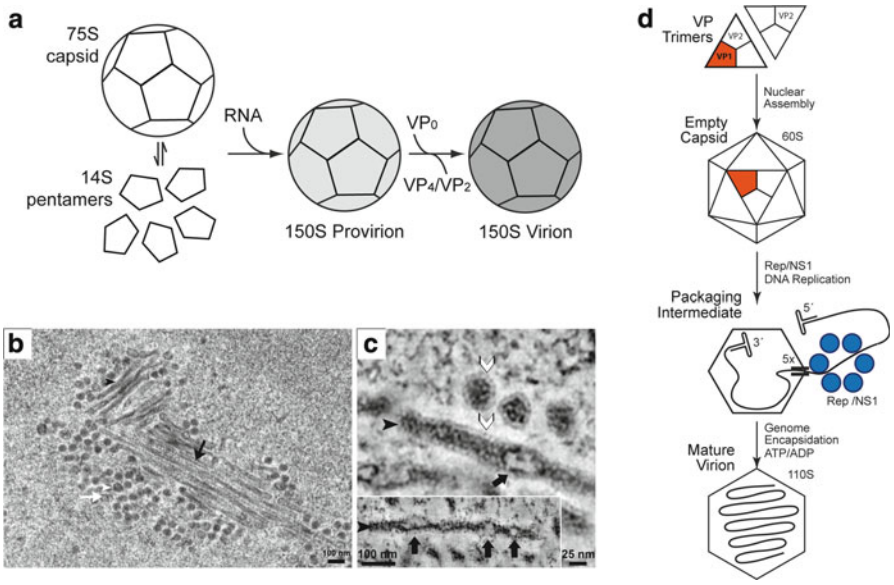


Fig. 10.5 Genome pakaging and virus maturation. (**a**) Late stages of Poliovirus assembly and virion maturation cleavage [30]. (**b**) Organization of the Polyomavirus factory. Electron microscopy of plastic-embedded thin sections of the nucleus of mouse fibroblasts showing tubular structures adjacent to a virus cluster (*black arrowhead*, "full" tubular structure; *white arrowhead*, full virion; *black arrow*, "empty" tubular structure; *white arrow*, empty virion). (**c**) Spherical virions and tubular structures seen in the nucleus of Polyomavirus-infected cells showing a lighter and well-arranged density at their periphery (*white arrowhead*) corresponding to capsid protein density, and a dense core (*black arrowhead*) suggesting DNA. (**b**, **c**: Adapted from [41]). (**d**) Integrative model of major stages in parvovirus assembly and maturation: I. VP capsid subunits assemble into trimers in the cytoplasm (*top image*). II. Inmature empty capsids (60S) form in the nucleus (*top center*). III. Genome packaging reaction (*bottom center*). IV. Mature viral particle (110S) (*bottom image*)

virus structure is known to atomic resolution, how the virions mature in the nucleus during productive infections is only partly understood. In a stepwise model, the capsid proteins would be sequentially added to and arranged on the viral DNA, resulting in its condensation and packaging to form the virion. The process implies multiple molecular recognition events of the viral genome by the CP subunits, in which DNA sequences near the viral origin of replication and the T-antigen (major replicating non-structural Polyomavirus protein) play important functions. Several DNA-binding domains (DBD) localized in all three VP proteins contribute to the packaging process. Major DBD of VP2 and VP3 were found in their C-terminal segments, whereas the DBD of VP1 was localized overlapping with the N-terminal bipartite NLS [42]. This VP1 N-terminal sequence of 15 amino acids is not visible in the virion crystal due to structural disorder, but it likely extends into the virion core to interact with the viral DNA.

The dynamics of maturation of the Polyomavirus and SV40 virions is not precisely known yet, but most evidences suggest a polymerization of the CP subunits onto the viral genome, acting as a scaffold. Although the intermediates are not well defined, the final nuclear assembly process seems to proceed through large, polymorphic structures that may serve as virus factories (Fig. 10.5b, c) (see Chap. 14 for a description of virus factories). Electron microscopy reveals tubular structures in the nucleus adjacent to clusters of virions. They share an organization consisting of a protein shell surrounding an electron dense DNA core, suggesting that they are the main factories able to yield mature virions budding from their ends, although the resolution mechanism is unknown.

10.5.3 Viral Genome Encapsidation in a Pre-formed Parvovirus Capsid

The final stages of parvovirus virions assembly and maturation are a subject of current active research, but in the best known cases it proceeds through the formation of large amounts of empty capsids in the nucleus, which serve as preformed substrate for active genome encapsidation. In MVM for example, empty capsids are first detected accumulated in the nucleus of synchronized cells preceding the accumulation of DNA-filled virions. A consensus model for parvovirus maturation, based on the available current data obtained in the AAV and MVM systems [43, 44] is outlined in Fig. 10.5d, although many details of the ssDNA packaging mechanism are still poorly understood. The reaction proceeds through a "packaging stage" in which the major non-structural proteins, NS1 in the *Parvovirus* genus and Rep 78/52 in AAV, mediate the association of the empty capsid with DNA replicative intermediates. These multifunctional proteins harbor helicase, DNA binding, and endonuclease activities that are critical for genome replication, and form oligomers *in vitro* in the presence of ATP. A Rep/NS1 hexameric packaging motor is probably formed at the five-fold symmetry axes, injecting the ssDNA genome through the channel of the capsid, which acts as the portal for genome encapsidation. For the packaging reaction the helicase activity of the non-structural proteins is essential. Genome encapsidation operates in the 3′ to 5′direction, mediated by specific contacts with the resolving hairpins and ATP hydrolysis. Encapsidation is coupled to DNA replication, ending as the capsid is filled with the complete ssDNA genome, which establishes regular contacts with some internal capsid residues and become icosahedrally ordered along certain sequences. The Rep/NS1 subunits may remain bound to the genome outside of the viral particle. The reader is referred to Chap. 12 for comparisons with equivalent mechanisms of dsDNA and dsRNA packaging into preformed bacteriophage capsids.

10.6 Perspectives and Conclusions

The assembly of simple icosahedral viruses proceeds through elegant genetically encoded and virus-specific pathways, which direct the self-association of the synthesized asymmetric protein subunits into a symmetric viral capsid shell. The general mechanism of capsid assembly involves first the oligomerization of CP subunits into early stable intermediates, usually of one major type per virus, which constitute the building blocks (CBBs). Capsid assembly then proceeds along a pathway regulated by ordered interactions between activated CBBs, most likely through a cascade of second-order reactions (see Chap. 19). The general architecture and the organization of the capsid subunits in the intermediates resembles those in the final capsid. However, the assembly process implies further structural rearrangements within and between the capsid subunits, including the establishment of multiple inter-protein non-covalent interactions, protein cleavages and covalent bonds (in some cases), and intracellular traffic of the assembly intermediates, prior to their final condensation into a closed container. Assembly is a dynamic process, implying that many of these interactions among subunits may have a transient morphogenetic role, and will not be preserved in the assembled virion. Functional capsid assembly cannot be understood without concomitant or later packaging of the genome, a complex process that viruses solve through molecular recognition patterns between the replicating genomes and either the assembly intermediates or a preformed empty capsid (see Chap. 12). During the final stages of morphogenesis of many viruses, cleavage of the viral particle leads to a mature infectious virion (see Chap. 13 for a description of maturation strategies in different viruses, including structurally complex bacteriophages). Virion morphogenesis in the cell, even for simple viruses, is frequently mediated by additional viral and or cellular factors, and may involve specific cellular factories (see Chap. 14). The efficiency of assembly critically determines some important features of viral fitness, such as the virus yield per host cell, or the acquisition of enough stability to prevail in natural enviroments.

The intensive research on small icosahedral viruses over the past decades have provided detailed information on the atomic structure of viral particles, and on the overall organization and mechanisms of assembly *in vitro* and in the host cells. Despite these advances, our knowledge on some crucial steps of the assembly of even structurally simple viruses is still poor or unclear in many respects, as exemplified by the following unsolved problems: the detailed structure of assembly intermediates in picornaviruses and parvoviruses; the cellular factors and transporters regulating VP traffic in the cell; the enzimology of genome encapsidation in parvoviruses; the processes of protein-protein and protein-nucleic acid recognition leading to the maturation of Poliovirus and Polyomavirus; or the resolution of the mature virions from the Polyomavirus factories. Ongoing research on these and other subjects related with morphogenesis promises exciting insights into essential aspects of the structural biology of simple icosahedral viruses in the near future.

Acknowledgements I gratefully acknowledge the collaborative work of Mavis Agbandje-Mckenna (University of Florida) on solving parvovirus MVMp capsid structure to atomic resolution, of Mauricio G. Mateu (CBMSO, Universidad Autónoma de Madrid) on MVM assembly, and of Thomas L. Benjamin (Harvard Medical School, Massachussets) on aspects related to Polyomavirus assembly. The many inspiring conversations along the years with regular participants to the "Virus Assembly" FASEB meeting (Vermont Academy, USA), in particular with Peter Tattersall (Yale University, Connecticut), are also deeply acknowledged. This review stands from the experience gained by the author on Parvovirus assembly thanks to the enthusiastic work of Eva Hernando, Eleuterio Lombardo, Beatriz Maroto, Juan C. Ramírez, Laura Riolobos, and Noelia Valle. Support on figures design by Jon Gil-Ranedo and Jorge Sánchez is also acknowledged. Research in JM Almendral's laboratory is currently supported by grant SAF 2011-29403 from the Spanish Ministerio de Ciencia e Innovación.

References and Further Reading

1. Agbandje-McKenna M, Llamas-Saiz AL, Wang F, Tattersall P, Rossmann MG (1998) Functional implications of the structure of the murine parvovirus, minute virus of mice. Structure 6:1369–1381
2. Kontou M, Govindasamy L, Nam HJ, Bryant N, Llamas-Saiz AL, Foces-Foces C, Hernando E, Rubio MP, McKenna R, Almendral JM, Agbandje-McKenna M (2005) Structural determinants of tissue tropism and *in vivo* pathogenicity for the parvovirus minute virus of mice. J Virol 79:10931–10943
3. Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ, Johnson JE, Kamer G, Luo M, Mosser AG, Ruecker RR (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 307:145–153
4. Hogle JM, Chow M, Filman DJ (1985) Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358–1365
5. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC (1991) Structure of simian virus 40 at 3.8-A resolution. Nature 28:278–284
6. Natarajan P, Lander GC, Shepherd CM, Reddy VS, Brooks CL III, Johnson JE (2005) Exploring icosahedral virus structures with VIPER. Nat Rev Microbiol 3:809–817
7. Tsao J, Chapman MS, Agbandje M, Keller W, Smith K, Wu H, Luo M, Smith TJ, Rossmann MG, Compans RW, Parrish CR (1991) The three-dimensional structure of canine parvovirus and its functional implications. Science 251:1456–1464
8. Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A, Chapman MS (2002) The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci U S A 99:10405–10410
9. Gurda BL, Parent KN, Bladek H, Sinkovits RS, DiMattia MA, Rence C, Castro A, McKenna R, Olson N, Brown K, Baker TS, Agbandje-McKenna M (2010) Human bocavirus capsid structure: insights into the structural repertoire of the parvoviridae. J Virol 84:5880–5889
10. Kaufmann B, Simpson AA, Rossmann MG (2004) The structure of human parvovirus B19. Proc Natl Acad Sci U S A 101:11628–11633
11. Racaniello VR (2001) Picornaviridae: the viruses and their replication. In: Knipe D, Howley P (eds) Fields virology. Lippincot Williams & Sons, New York, pp 685–722
12. Tullis GE, Lisa RB, Pintel DJ (1993) The minor capsid protein VP1 of the autonomous parvovirus minute virus of mice is dispensable for encapsidation of progeny single-stranded DNA but is required for infectivity. J Virol 67:131–141
13. Zadori Z, Szelei J, Lacoste MC, Li Y, Gariepy S, Raymond P, Allaire M, Nabi IR, Tijssen P (2001) A viral phospholipase A2 is required for parvovirus infectivity. Dev Cell 1:291–302

14. Lombardo E, Ramírez JC, García J, Almendral JM (2002) Complementary roles of multiple nuclear targeting signals in the capsid proteins of the parvovirus minute virus of mice during assembly and onset of infection. J Virol 76:7049–7059
15. Lombardo E, Ramirez JC, Agbandje-McKenna M, Almendral JM (2000) A beta-stranded motif drives capsid protein oligomers of the parvovirus minute virus of mice into the nucleus for viral assembly. J Virol 74:3804–3814
16. Riolobos L, Reguera J, Mateu MG, Almendral JM (2006) Nuclear transport of trimeric assembly intermediates exerts a morphogenetic control on the icosahedral parvovirus capsid. J Mol Biol 357:1026–1038
17. Chen XS, Stehle T, Harrison SC (1998) Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. EMBO J 17:3233–3240
18. Reguera J, Carreira A, Riolobos L, Almendral JM, Mateu MG (2004) Role of interfacial amino acid residues in assembly, stability, and conformation of a spherical virus capsid. Proc Natl Acad Sci USA 101:2724–2729
19. Li PP, Nakanishi A, Clark SW, Kasamatsu H (2002) Formation of transitory intrachain and interchain disulfide bonds accompanies the folding and oligomerization of simian virus 40 Vp1 in the cytoplasm. Proc Natl Acad Sci USA 99:1353–1358
20. Raices M, D'Angelo MA (2012) Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. Nat Rev Mol Cell Biol 13:687–699
21. Robbins J, Dilworth SM, Laskey RA, Dingwall C (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64:615–623
22. Mattaj IW, Englmeier L (1998) Nucleocytoplasmic transport: the soluble phase. Annu Rev Biochem 67:265–306
23. Chang D, Haynes JI, Brady JN, Cosigli RA (1992) Identification of a nuclear localization sequence in the polyomavirus capsid protein VP2. Virology 191:978–983
24. Moreland RB, Garcea RL (1991) Characterization of a nuclear localization sequence in the polyomavirus capsid protein VP1. Virology 185:513–518
25. Ishii N, Minami N, Chen EY, Medina AL, Chico MM, Kasamatsu H (1996) Analysis of a nuclear localization signal of simian virus 40 major capsid protein Vp1. J Virol 70:1317–1322
26. Ishii N, Nakanishi A, Yamada M, Macalalad MH, Kasamatsu H (1994) Functional complementation of nuclear targeting-detective mutants of simian virus 40 structural proteins. J Virol 68:8209–8216
27. Vihinen-Ranta M, Wang D, Weichert WS, Parrish CR (2002) The VP1 N-terminal sequence of canine parvovirus affects nuclear transport of capsids and efficient cell infection. J Virol 76:1884–1891
28. Grieger JC, Snowdy S, Samulski RJ (2006) Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly. J Virol 80:5199–5210
29. Johnson JE (1996) Functional implications of protein-protein interactions in incosahedral viruses. Proc Natl Acad Sci USA 93:27–33
30. Hogle JM (2002) Poliovirus cell entry: common structural themes in viral cell entry. Annu Rev Microbiol 56:677–702
31. Sonntag F, Schmidt K, Kleinschmidt JA (2010) A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A 107:10220–10225
32. Riolobos L, Valle N, Hernando E, Maroto B, Kann M, Almendral JM (2010) Viral oncolysis that targets Raf-1 signaling control of nuclear transport. J Virol 84:2090–2099
33. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol XXVII:1–24
34. Salunke DM, Caspar DL, Garcea RL (1986) Self-assembly of purified polyomavirus capsid protein VP1. Cell 46:895–904

35. Hernando E, Llamas-Saiz AL, Foces-Foces C, McKenna R, Portman L, Agbandje-McKenna M, Almendral JM (2000) Biochemical and physical characterization of parvovirus minute virus of mice virus-like particles. Virology 267:299–309
36. Yuan W, Parrish CR (2001) Canine parvovirus capsid assembly and differences in mammalian and insect cells. Virology 279:546–557
37. Maroto B, Ramírez JC, Almendral JM (2000) Phosphorylation status of the parvovirus minute virus of mice particle: mapping and biological relevance of the major phosphorylation sites. J Virol 74:10892–10902
38. Garcea R, Benjamin T-L (1983) Host range of trasnforming gene of polyoma virus plays a role in virus assembly. Proc Natl Acad Sci USA 80:3613–3617
39. Garcea RL, Ballmer-Hofer K, Benjamin TL (1985) Virion assembly defect of polyomavirus *hr-t* mutants: underphosphorylation of major capsid protein VP1 before viral DNA encapsidation. J Virol 54:311–316
40. Naumer M, Sonntag F, Schmidt K, Nieto K, Panke C, Davey NE, Popa-Wagner R, Kleinschmid JA (2012) Properties of the AAV assembly activating protein AAP. J Virol 86:13038–13048
41. Erickson KD, Bouchet-Marquis C, Heiser K, Szomolanyi-Tsuda E, Mishra R, Lamothe B, Hoenger A, Garcea RL (2012) Virion assembly factories in the nucleus of polyomavirus-infected cells. PLoS Pathog 8:1–15
42. Li PP, Nakanishi A, Shum D, Sun PC, Salazar AM, Fernandez CF, Chan S-W, Kasamatsu H (2001) Simian virus 40 VP1 DNA-binding domain is functionally separable form the overlapping nuclear localization signal and is required for effective virion formation and full viability. J Virol 75:7321–7329
43. King JA, Dubielzig R, Grimm SW, Kleinschmidt JA (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. EMBO J 20:3282–3291
44. Plevka P, Hafenstein S, Li L, D'Abramo A, Cotmore SF, Rossmann MG, Tattersall P (2011) Structure of a packaging-defective mutant of minute virus of mice indicates that the genome is packaged *via* a pore at a 5-fold axis. J Virol 85:4822–4827
45. Li M, Garcea RL (1994) Identification of threonine phosphorylation sites on the polymavirus major capsid protein VP-1: relationship to the activity of middle T antigen. J Virol 68:320–327

Further Reading

Carrasco L, Almendral JM (eds) (2006) Virus patógenos. Hélice and Fundación BBVA, Madrid
Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2009) Principles of virology, 3rd edn. ASM Press, Washington, DC
Kerr JK, Cotmore SF, Bloom ME, Linden RM, Parrish CR (eds) (2006) Parvoviruses. Hodder Arnold, London
Harrison SC (2006) Principles of virus structure. In: Knipe DM, Howley PM (eds in chief) Fields virology, 5th edn. Lipincott Williams and Wilkins, Philadelphia, p 59–98
Reddy VS, Johnson JE (2005) Structure-derived insights into virus assembly. Adv Virus Res 64:45–68
Zlotnick A, Fane BA (1997) Mechanisms of icosahedral virus assembly. In: Agbandje-McKenna M, McKenna R (eds) Structural biology. RSC Publishing, Cambridge, pp 180–202

Chapter 11
Structure and Assembly of Complex Viruses

Carmen San Martín

Abstract Viral particles consist essentially of a proteinaceous capsid protecting a genome and involved also in many functions during the virus life cycle. In simple viruses, the capsid consists of a number of copies of the same, or a few different proteins organized into a symmetric oligomer. Structurally complex viruses present a larger variety of components in their capsids than simple viruses. They may contain accessory proteins with specific architectural or functional roles; or incorporate non-proteic elements such as lipids. They present a range of geometrical variability, from slight deviations from the icosahedral symmetry to complete asymmetry or even pleomorphism. Putting together the many different elements in the virion requires an extra effort to achieve correct assembly, and thus complex viruses require sophisticated mechanisms to regulate morphogenesis. This chapter provides a general view of the structure and assembly of complex viruses.

Keywords Virus structure • Virus assembly • Symmetry • Capsid • Cementing proteins • Envelope • Symmetry mismatch • Scaffold • Maturation • Virus evolution

Abbreviations

3D	Three-dimensional
ABV	*Acidianus* bottle-shaped virus
ATV	*Acidianus* two-tailed virus
cryo-EM	Cryo-electron microscopy
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA

C. San Martín (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: carmen@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_11,

EMDB	Electron Microscopy Data Bank
GON	Group of nine
GOS	Group of six
PBCV-1	*Paramecium bursaria Chlorella* virus-1
PDB	Protein Data Bank
Sid	Size determination protein
SNDV	*Sulfolobus neozelandicus* droplet-shaped virus
ssDNA	Single-stranded DNA
SSIP-1	*Salisaeta* icosahedral phage 1
ssRNA	Single-stranded RNA
STIV	*Sulfolobus* turreted icosahedral virus

11.1 Introduction

A viral particle consists essentially of a proteinaceous capsid with multiple roles in protection of the viral genome, cell recognition and entry, intracellular trafficking and controlled uncoating. Evolutionary forces have caused viruses to adopt different strategies to achieve these goals. Simple viruses (Chap. 10) generally build their capsids from a number of copies of the same, or a few different proteins, organized into a symmetric oligomer. In the case of complex viruses, capsid assembly requires further elaborations. What are the main characteristics that define a structurally complex virus?

Structural complexity on a virus often, but not necessarily, derives from the need to house a large genome, in which case a larger capsid is required. However, capsid or genome sizes by themselves are not determinants of complexity. For example, flexible filamentous viruses can reach lengths in the order of microns, but most of their capsid mass is built by a single capsid protein arranged in a helical pattern [1]. On the other hand, architecturally complex viruses such as HIV have moderate sized genomes (7–10 kb of single-stranded (ss) RNA) [2]. Structurally complex viruses incorporate a larger variety of components into their capsids than simple viruses. They may contain accessory proteins with specific architectural or functional roles; or incorporate non-proteic elements such as lipids.

The elaborated composition of complex virus particles often involves a rupture of the basic symmetry rules (Chap. 2), from a range of symmetry mismatches in icosahedral shells, to completely asymmetric or pleomorphic capsids. The more subtle departure from symmetry is the case when identical subunits occupy similar but slightly different environments, as in the case of quasi-equivalence in icosahedral shells (see Chap. 2). In other cases, virion components with different symmetry may interact with each other, forming a *symmetry mismatch* at the interface. In the extreme case, identical components may form morphological units with no

symmetry, or even assemble in a completely different manner for each realization of the virion – this property is called *pleomorphism*. Recent advances in cryo-electron microscopy (cryo-EM) (Chap. 3) and X-ray crystallography (Chap. 4) are helping to unveil the organization of complex viruses in great detail, including features that depart from strict icosahedral symmetry. Understanding the architectural details of asymmetric capsids is the most challenging problem, since structural biology techniques heavily rely on the use of symmetries to reach high resolution detail. Cryo-electron tomography (Chap. 3) is helping to advance our understanding of these viruses, although the resolution currently attained is still in the 3–5 nm range.

The presence of many different elements in the virion entails an extra effort to achieve correct assembly. Accordingly, complex virus morphogenesis requires sophisticated mechanisms, tightly regulated in space and time. Here we provide a general view of all these variations in complexity, finishing with a consideration on the evolutionary insights provided by structural studies on complex viruses.

11.2 Molecular Composition of Complex Viruses

11.2.1 Different Proteins with Specific Roles

A characteristic feature of complex viruses is the presence of multiple proteins in the virion, playing specific architectural or functional roles during the viral cycle. For example, different proteins may occupy the sixfold and fivefold coordinated positions in the icosahedral net (see Chap. 2). The specific architectural role of proteins at the fivefold vertices is often combined with a specific functional role, as will be described in Sect. 11.3. In icosahedral viruses with triangulation numbers $T > 1$, mobile terminal regions of the capsid proteins may adopt different conformations depending on their position in the capsid. In this way they act as molecular switches, enabling the same protein to occupy the different quasi-equivalent environments (see Chap. 2). In complex viruses, these mobile arms may still exist, but they often appear combined with a variety of minor capsid proteins, required for correct assembly of the virion. These *cementing proteins* can be considered as detached molecular switches, required to modulate the variety of interactions needed for assembly and stability of a complex capsid. One case where the intricate capsid organization includes: (i) biochemically different hexameric and pentameric capsomers; (ii) a network of mobile arms; and (iii) cementing polypeptides, has recently been described in great detail: adenovirus [3, 4].

Adenoviruses infect vertebrates. They enclose their dsDNA genome within a *pseudo* $T = 25$ icosahedral shell with a diameter of 950 Å between vertices. Trimers of the major capsid protein, hexon, constitute all the sixfold capsomers, while the vertices are occupied by pentamers of a protein named penton base. Although there is no sequence similarity between hexon and penton base, both have

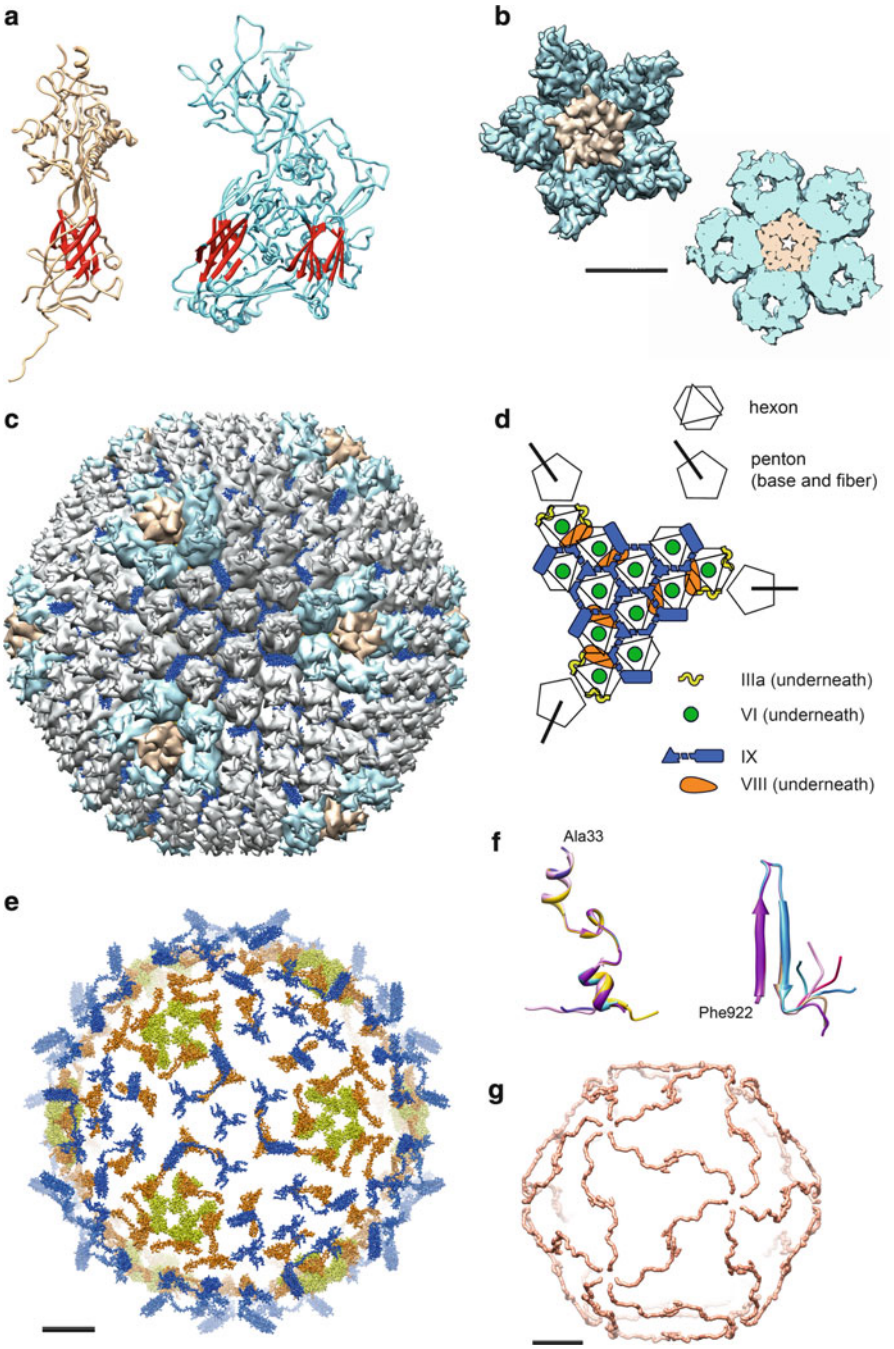


Fig. 11.1 Molecular composition of complex viruses: capsid proteins, cementing proteins, and mobile arms. (**a**) Structure of adenovirus penton base (*beige*) and hexon (*cyan*) monomers, with

a very similar fold based on the eight-stranded β-barrel motif, or "jelly roll", which allows appropriate interdigitation of the different capsomers to form the closed shell [5, 6]. Penton base folds as a single jelly roll. The hexon monomer folds as a double β-barrel jelly-roll perpendicular to the capsid surface (Fig. 11.1a). Thanks to this fold, trimers have a pseudo-hexagonal shape and can occupy sixfold symmetry positions in the icosahedron. Each capsid facet is formed by 12 trimers of hexon. The general icosahedral architecture of adenovirus can be described as two different systems of tiles. Nine hexon trimers form the central plate of each facet, known as GON (Group of Nine). The five peripentonal hexon trimers, together with the penton base, form the second tile system, known as GOS (Group of Six) (Fig. 11.1b). Numerous double-stranded (ds) DNA viruses share the basic jelly-roll fold in their capsid proteins and assemble capsids organized as combinations of tiles based on pentagonal and pseudo-hexagonal building blocks. In larger viruses, the tiles composed by the pentameric vertex and surrounding capsomers are known as *pentasymmetrons*, while the triangular tiles centered at the threefold icosahedral axis are called *trisymmetrons*. This efficient architectural solution facilitates construction of very large icosahedral capsids; triangulation numbers as large as $T = 169$ have been described, and larger ones are likely to exist [8, 9].

In adenovirus, apart from hexon and penton base, there are at least four other proteins making up the icosahedral shell. Minor capsid proteins IIIa, VI, VIII and IX are required for correct capsid assembly and occupy specific positions in the capsid (Figs. 11.1c, d, e), forming specialized networks that stabilize the two systems of tiles [3]. Polypeptide IX is the only cementing protein located on the outer part of the adenovirus capsid. It has an extended structure and forms a sort of hairnet keeping together the hexon trimers in each GON, and binding GONs to

Fig. 11.1 (continued) the β-barrel jelly roll motifs highlighted in *red*. The molecules are oriented so that the external capsid surface is up. Notice the extended N-terminal arm in penton base directed towards the interior of the virion (Unless otherwise indicated, all ribbon and surface structure representations in this chapter, as well as fullerene models, were prepared using USCF Chimera software [7]). (**b**) One penton base pentamer and the five peripentonal hexon trimers form the GOS, the adenovirus pentasymmetron. *Left*, view of the GOS from outside the capsid. *Right*, a slab showing the tight interdigitation at the base of the capsomers to close the shell. Color scheme as in (**a**). (**c**) Adenovirus capsid, seen along a threefold icosahedral axis. The GOS is colored as in (**b**), and the rest of the hexon trimers are in *gray*. These form the central plate of the icosahedral facet, termed GON, or trisymmetron in the more general terminology. In *blue*, the cementing protein polypeptide IX. (**d**) Schematics showing the location of cementing proteins in an adenovirus capsid facet. Reproduced from [4]. (**e**) Cementing proteins in the adenovirus capsid. The view is as in (**c**), with the pentons and hexons removed to reveal both the external and internal networks of accessory proteins. These are colored as in (**d**). Notice that polypeptide VI has not been traced in the high resolution structure. (**f**) Mobile arms in adenovirus hexon. Different conformations of the hexon N-terminal (*left*) and C-terminal (*right*) regions in the capsid are shown. Panels (a), (b), (c), (e) and (f) made from atomic coordinates deposited at the Protein Data Bank (PDB; Protein Data Bank is at http://www.pdb.org) with entry ID 3IYN. (**g**) Minor capsid proteins as size determinants. The bacteriophage PRD1 "tape measure" protein P30 forms a cage beneath the capsid surface (PDB ID 1W8X). Scale bars correspond to 100 Å

GONs across the icosahedral edges. The N-terminal domains of three IX monomers join *via* hydrophobic interactions at the icosahedral and local threefold axes in the GONs forming triskelion structures. A long, unstructured domain of each monomer runs in a different direction towards the facet edges, where the C-terminal α-helix joins with the C-terminal helices of another three copies of IX, different from those forming the N-terminal triskelion, to create a leucine 4-helix bundle. On the interior of the shell, each GON is further stabilized by copies of polypeptide VIII located around the icosahedral threefold symmetry axis. Also on the inner capsid surface, polypeptide IIIa mediates the interaction between penton base and the peripentonal hexons, to keep each GOS together. Finally, IIIa and VIII cooperate to bind each GOS to its five surrounding GONs. The remaining minor capsid protein, polypeptide VI, has not been unequivocally traced so far, but has been assigned to density within an internal cavity present in each hexon trimer.

Mobile regions of hexon and penton base also play a role in the extensive interaction networks in the adenovirus capsid. Due to their flexibility, these regions could not be traced in the crystal structures of the isolated proteins, but they adopt ordered conformations when they are within the capsid context. The N- and C-termini of the hexon monomer, located at the innermost part of the capsomer, adopt a total of 5 (N-) and 6 (C-) different conformations to establish interactions between neighbouring hexons, or between hexons and minor capsid proteins (Fig. 11.1f). Similarly, for each penton base monomer an N-terminal arm extends away from the β-barrels that form the main body of the protein towards the viral core, interacting with two IIIa monomers along the way, and therefore contributing to anchor the penton within the GOS. Interestingly, some of the interactions between cementing proteins and hexons, and among cementing proteins, occur by β-sheet augmentation. That is, the interaction is mediated by a β-strand from one of the proteins binding to the edge of a β-sheet in the other. This observation tells about the intricate organization of the capsid and makes us wonder about the difficulty of assembling all elements together. The fact that no high resolution structure is available for any of the minor capsid proteins in isolation suggests that they may require the virion context to fold properly.

Although in general it is understood that cementing proteins are required for correct viral assembly, it is difficult to pinpoint their exact role in morphogenesis. Some of them are dispensable for assembly, but required to reach structural stability; this is the case of adenovirus polypeptide IX [10]. Others are thought to play the role of "molecular rulers", determining capsid size. This role was proposed when the structure of bacteriophage PRD1 was solved by protein crystallography (see also Chap. 17). It was then found that minor capsid protein P30, required for assembly, runs beneath the icosahedral edges, from the vertex to the twofold symmetry axis. Thanks to its extended conformation, two copies of the 83-residue polypeptide can cover the 300 Å length of the capsid edge, and act as a tape measure during morphogenesis [11] (Fig. 11.1g). A further complication for determining the role of minor virion components in assembly comes from the fact that, in keeping with the genetic economy principle, they often play other roles different from the purely architectural one during the viral cycle. A remarkable example of this phenomenon is illustrated by adenovirus polypeptide VI [4]. This protein is

involved in disrupting the endosomal membrane, so that the virus can escape into the cytosol after internalization. It also has a role in facilitating virion traffic to the nucleus along the microtubular network; acts as an activator of the adenoviral gene expression; and promotes transport of newly synthesized hexon to the nucleus. Finally, a C-terminal peptide of polypeptide VI activates the viral protease for maturation.

Other additional proteins may be incorporated to the viral particles and play fundamental roles for viability. Elucidating the organization of these additional components within the virion is not straightforward, since they usually do not follow defined symmetry rules, and their disposition may even change between particles of the same virus. Notable examples are viral proteases, such as the maturation protease VP24 in herpesvirus [12], or the adenoviral protease AVP [4]; and molecular motors involved in nucleic acid translocation, such as dsDNA packaging ATPases in bacteriophage and herpesviruses [13, 14], or the dsRNA packaging ATPase in cystoviruses (bacteriophage Φ6) [15]. Viruses with RNA genomes must carry their own replication and transcription enzymes, to supply RNA metabolism functions absent in the cell [16]. Viruses that carry out their replication in the cytosol (*e.g.*, vaccinia) must also supply DNA and RNA processing enzymes whose cellular counterparts are only present in the nucleus [17]. Some dsDNA viruses encapsidate basic proteins that help screen the nucleic acid charge repulsion, to facilitate compaction of the genome within the reduced capsid space. These basic proteins can be of cellular origin, such as in Simian Virus 40, which uses histones to pack its minichromosome [18]. Baculovirus [19], adenovirus [20], mimivirus [21] and poxviruses [17] encode their own DNA compacting proteins. The genomes of negative strand ssRNA viruses usually appear in the form of ribnucleoproteic structures [22]. More information on the packaging motors and on the organization of nucleic acids within viral capsids can be found in Chap. 12.

11.2.2 Membranes

Apart from the genome and structural proteins, a large number of viruses incorporate lipidic layers into their architecture. Lipid bilayers (membranes) are widely extended in biological entities such as cells and organelles, and are ideally suited for enclosing a defined volume and separating it from neighboring compartments or the surrounding environment. This is the same function they play in viral capsids. For viruses, membranes are particularly advantageous, since they can readily be taken from the cell, are highly scalable in size, and do not consume coding space in the genome. More detail on how viruses sequester cell membranes for their own use can be found in Chap. 14.

Viruses with external membranes are called *enveloped viruses*. Envelopes form a protective layer, blocking entry of aggressive chemicals or enzymes into the viral particle. They are composed not only by lipids but also by protein and sugars

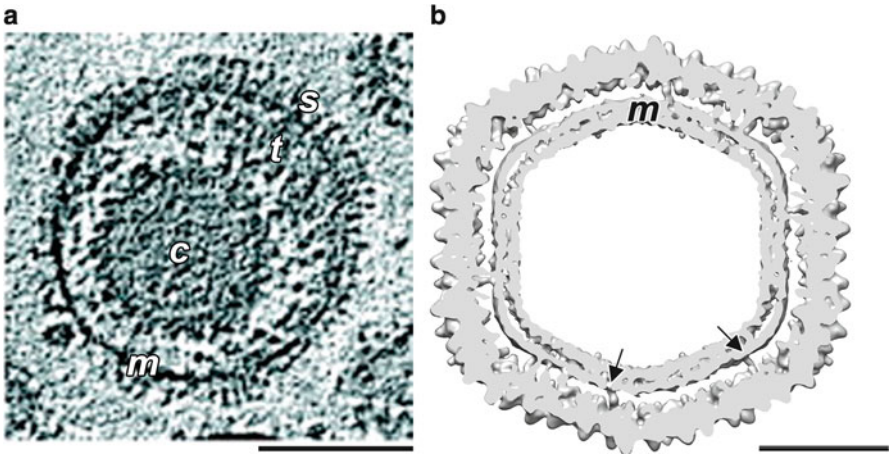


Fig. 11.2 Viruses with lipid bilayers. (**a**) Slice of a cryo-electron tomography reconstruction of herpes simplex virus-1 showing the icosahedral capsid (*c*), tegument (*t*), membrane (*m*) and glycoprotein spikes (*s*). Scale bar, 1,000 Å (Modified from [25]). (**b**) Bacteriophage PRD1 has an internal membrane (*m*) that adopts an icosahedral profile due to interactions with capsid proteins (*arrows*). Shown is a slab of the cryo-EM map deposited at the Electron Microscopy Data Bank (EMDB; Electron Microscopy Data Bank is at http://www.ebi.ac.uk/pdbe/emdb/) with ID EMD_1013, viewed along a twofold icosahedral axis. The scale bar corresponds to 100 Å

(glycoproteins). An implication of this arrangement is that in enveloped viruses infectivity is linked to membrane integrity, because the viral attachment proteins which recognize the host cell receptors and trigger internalization are in the viral envelope. Such is the case of influenza virus hemaglutinins [23], or Env proteins in retroviruses [2]. In the simpler enveloped viruses, such as Alphaviruses, the glycoproteins completely cover the lipid surface and follow the same icosahedral organization as the inner shell (T = 4) [24]. In other cases, such as retroviruses or herpesviruses (Fig. 11.2a), the structure directly in contact with the genome (nucleocapsid or capsid) is not in contact with the membrane, and the organization of glycoprotein spikes is irregular and does not reflect the organization of the virion inner contents [25, 26].

The membrane can also be located beneath the icosahedral shell, such as in Tectiviruses (PRD1) and structurally related viruses [11, 24]. In these cases, the membrane itself adopts an icosahedral layout forced by interactions with the capsid proteins (Fig. 11.2b). In PRD1, some of the mobile arms in its major capsid protein are embedded in the outer leaflet of the membrane, contributing to enforce the icosahedral shape [11]. Icosahedral viruses with internal membranes have a large complement of virion-encoded membrane proteins: approximately half of the 18 proteins present in the PRD1 virion are membrane proteins [27]. The membrane in PRD1 can undergo a large conformational change and protrude forming a tube from one of the vertices. It has been proposed that this tube has a function in injecting the

viral genome into the host cell, similar to that played by tail structures in other bacteriophages [28] (see Chap. 17).

Some of the most architecturally complex viruses are enveloped. For example, herpesviruses have a T = 16 icosahedral capsid (1,500 Å in diameter), formed by one major capsid protein and several accessory proteins. This capsid is surrounded by a thick tegument layer, containing at least 13 different viral proteins and also some cellular components. Capsid and tegument are enveloped by a membrane with more than 12 different types of viral glycoproteins [12, 25]. Large dsDNA viruses infecting aquatic eukaryotic microorganisms, including the giant *Acanthamoeba polyphaga* Mimivirus (diameter ~750 nm), have internal membranes like bacteriophage PRD1 [21, 29]. Asfarviruses (African swine fever virus, diameter ~200 nm) have both an internal membrane surrounded by an icosahedral shell and a loose external envelope [30]. Other examples of complex, lipid-containing viruses will be described in Sect. 11.4.

11.3 Departures from Symmetry in Quasi-Icosahedral Capsids

11.3.1 Layers with Different T Numbers

In Sect. 11.2.2, it was pointed out that viruses could be organized in multiple layers, intercalating protein (ordered or not) and lipids. In other cases such as adenovirus, multiple cementing proteins combine to form a single icosahedral capsid, while additional components (*e.g.*, dsDNA condensing proteins) do not show any symmetrical organization. In yet another instance, concentric icosahedral protein shells are formed. Remarkably, these shells may have different triangulation numbers, including some not predicted by the theory of quasi-equivalence. This type of organization is most prominently present in the dsRNA Reoviruses.

Rotaviruses and mammalian orthoreoviruses are the best characterized members of the *Reoviridae* family [16]. The mature rotavirus virion has a diameter of approximately 100 nm, and is organized in three concentric layers composed by four different proteins (Fig. 11.3). The innermost layer, or core shell, is formed by 120 molecules of protein VP2. Because of the number of protein monomers in this shell, it is described as having a T = 2 triangulation number, a conformation not allowed in the Caspar and Klug formalism [31] (Chap. 2). In fact, the VP2 monomers adopt two different conformations to yield a T = 1 icosahedron of asymmetric dimers. VP2 works as a platform for assembly of the next layer, formed by 260 trimers of protein VP6 in a T = 13 lattice. VP6 must be able to adopt not only the different conformations required for the T = 13 quasi-equivalence requirements, but also those extra ones required to compensate for the symmetry mismatch with the VP2 layer. The third, and outermost layer, is also a T = 13 icosahedron composed by 260 trimers of glycoprotein VP7, plus 60 spikes formed by trimers of VP4 that protrude from peripentonal positions. The VP4 spikes must

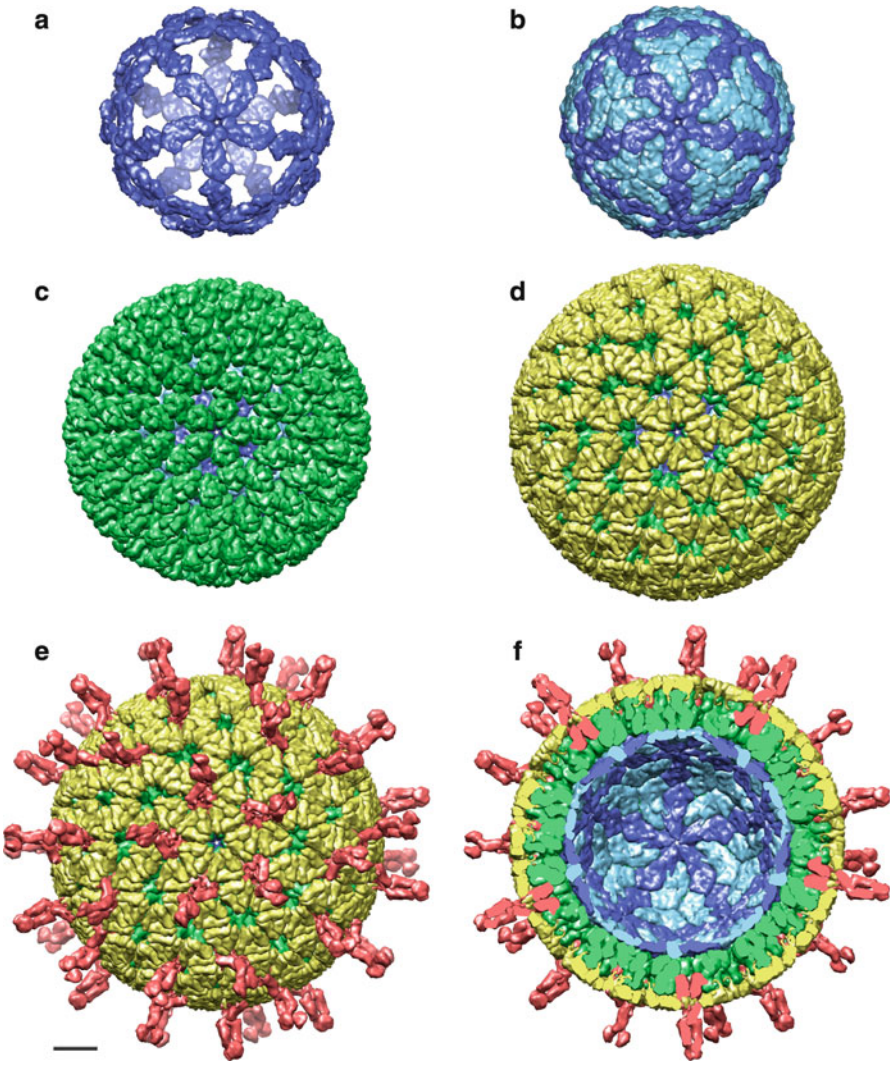


Fig. 11.3 Multiple layered viruses: rotavirus capsid architecture. Panels (**a**) to (**e**) show the consecutive building layers of the mature rotavirus virion. (**a**) Only one type of VP2 monomer in the core shell is represented. (**b**) The VP2 core shell, a T = 1 particle formed by asymmetric dimers. (**c**) The intermediate T = 13 VP6 layer. (**d**) VP7 glycoprotein layer. (**e**) The complete virion with VP4 spikes. (**f**) Cut out representation of the virion depicting the layered organization. The view is along a fivefold icosahedral symmetry axis. The scale bar represents 100 Å. Figure prepared using atomic coordinates from PDB IDs 3N09 and 3IYU

be proteolytically cleaved for the virus to be infectious. Remarkably, after cleavage the fragments remain non-covalently associated on the virion surface, but undergo an intriguing conformational modification, changing from a trimeric arrangement to

a mixture of trimeric, dimeric and monomeric associations. The VP1 and VP3 transcription enzymes are also part of the virion, and are located beneath the core shell surface.

The orthoreovirus capsid is also triple layered, and shares architectural similarities with rotavirus, namely the mismatch between "T = 2" and T = 13 symmetries. However, the composition is more complex, with six different proteins instead of four. The major differences appear in the outermost layers, probably reflecting differences in the viral entry mechanism. Instead of having 60 short spikes distributed in the icosahedral facet, orthoreovirus displays large turreted structures combined with a long, flexible fiber in each of the 12 vertices.

Why do viruses have different layers? As in the case of membranes, protein layers help to separate different compartments, and most likely also different functions along the infectious cycle. dsRNA viruses need to keep their genome confined within the core shell at all times during infection, to protect it from aggressions by cellular nucleases, and to prevent antiviral reactions triggered by accumulation of dsRNA. In rotavirus, the double layered particle formed by VP2 and VP6 is the transcriptionally competent form of the virus. The external layers carry the viral components in charge of initial interaction with the host: recognition, attachment and entry. These are shed once entry into a new host cell has been accomplished [32]. In reovirus however, the turrets are not lost upon entry like the other external layers, but form part of the double layered, transcriptionally active form of the virus [33].

11.3.2 Symmetry Mismatches

In the previous section, we have seen that in reoviruses there is a symmetry mismatch between two concentric shells with different triangulation numbers. Nevertheless, the two layers still follow icosahedral symmetry, and therefore it has been possible to study their organization at a very detailed level. Symmetry mismatches (two elements with different symmetries in direct interaction) are frequent in icosahedral viral structures, particularly at the vertices, where proteins involved in genome translocation or host attachment reside. Solving the organization of mismatched features represents a remarkable challenge for structural biology techniques, due to the predominance of icosahedral symmetry in the complete virion that obscures them. In the *Cystoviridae* representative bacteriophage Φ6, a hexameric ssRNA packaging ATPase occupies multiple fivefold vertices of the empty procapsid [15]. In dsDNA bacteriophages, 12-fold portal structures occupy one of the vertices (Sect. 11.3.4 and Chaps. 12 and 17). Host recognition elements often take the shape of elongated fibers protruding from the fivefold capsomers. The oligomerization state and number of fibers per vertex varies, and is usually at odds with the pentameric architecture of the capsomers.

In human adenoviruses, a trimer of the fiber forms a non-covalent complex with a pentamer of penton base (Fig. 11.4a). The C-terminal domain of fiber is the main

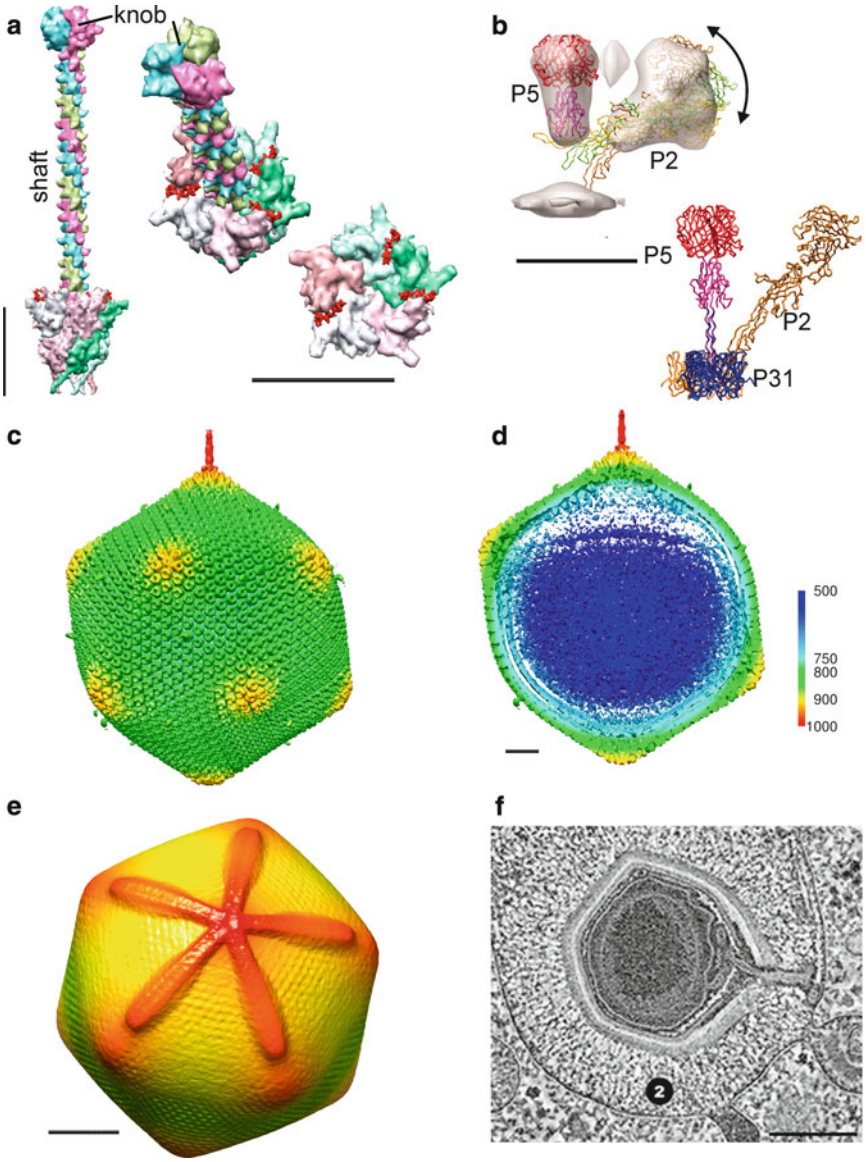


Fig. 11.4 Symmetry mismatches and singular vertices. (**a**) Adenovirus vertex complex, composed by pentameric penton base (compare with Fig. 1a and b) and the trimeric fiber (knob and distal shaft from PDB ID 1QIU, rest of the shaft modeled). In *red*, three N-terminal fiber peptides bound to penton base (PDB ID 3IZO). *Left*: *side view*; *center*, *oblique view*; *right*, *top view* of the penton base with the fiber shaft and knob removed to highlight the position of the N-terminal fiber peptides. (**b**) A model for the double fiber complex in bacteriophage PRD1 vertex. *Top*: high resolution structures of trimeric P5 and monomeric P2 fitted to a cryo-EM reconstruction of the vertex region (*side view*). The *arrow* indicates variability in the position of P2. Bottom: a model of the complex interactions between the three vertex components. P31 forms the fivefold capsomers

player in trimerization and forms the distal knob, responsible for binding to the adenovirus primary receptor. The knob is followed by a shaft of variable length and flexibility, depending on the serotype. The knob domain has an eight-stranded β-sandwich fold similar to that of the major capsid protein, and the shaft forms a triple β-spiral [36]. Finally, the N-terminal region is responsible for binding to penton base. At the proximal part the shaft becomes frayed and three flexible N-terminal peptides spread out to attach to the penton base pentamer. The crystal structure of penton base in complex with an N-terminal fiber peptide [5] showed the peptide bound to a groove on the outer surface of the pentamer formed by the interface between two penton base monomers, reaching radially from the penton center to the outer rim. Five peptides were observed with equivalent density, implying that all fiber binding sites are equivalent. Therefore, in the virus the three N-terminal tails may adopt two different arrangements: either they occupy three consecutive grooves, or two of them are in neighboring binding sites and the third one is flanked by two empty grooves. In the recently solved cryo-EM atomic structure of the complete virion, density for the proximal part of the fiber shaft was observed protruding from the center of the penton base pentamer [3, 37]. Density for the fiber shaft is blurred by the enforced fivefold symmetry, but it could be observed that its base interacts with a hydrophobic ring at the rim of a narrow channel in the center of the penton. This hydrophobic interaction may allow relative rotation of fiber on penton to accommodate the symmetry mismatch, while the N-terminal tails secure the binding to the penton groves. Fiber binding to the adenovirus receptor in the cell results in its release from the capsid, which in turn induces a cooperative conformational change in the penton base pentamer. This change is thought to play a role in preparing pentons for release at a later stage of adenovirus entry and uncoating [4].

In orthoreovirus, the receptor binding protein σ1 also forms a flexible trimeric spike attached to the fivefold turret [16, 38]. Intriguingly, some viruses can hold two fibers attached to the same vertex: this is the case of fowl adenovirus type-1, where two fibers of different length can be observed bound to the same penton base [39]. Bacteriophage PRD1 also has two different spikes (proteins P2 and P5) attached to a single pentamer of the vertex protein P31 (Fig. 11.4b). Moreover, in PRD1 each spike has a different oligomerization state. P5 is a trimer resembling the structure of the adenovirus fiber; while the other, P2, is a monomer with a pseudo-β propeller

Fig. 11.4 (continued) (pentons) in PRD1. This study showed that the two fibers interact with each other at the icosahedral capsid level, and that P2 can move relatively to P5 (Reproduced from [34]. With permission). (**c**) The PBCV-1 capsid (EMDB ID EMD_1597) showing the special vertex with a spike, and (**d**), a central slab where the asymmetry in internal contents can be appreciated. Color key indicates color changes with map radius. (**e**) 3D map of the Mimivirus capsid (EMDB ID EMD_5039) showing the starfish feature (*red*). (**f**) Slice of a tomographic reconstruction showing a Mimivirus particle within a phagosome. The viral membrane is extruded through the open stargate (Modified from [35]. Scale bars represent 100 Å in panels (**a**) and (**b**); 200 Å in (**c**) and (**d**); and 1,000 Å in (**e**) and (**f**)

head. The role of each spike in PRD1 host recognition and attachment is not fully clarified [27, 34] (see also Chap. 17).

The biological significance of these symmetry mismatches has long intrigued virologists. For the dodecameric nucleic acid packaging motors, the mismatch may allow conformational changes required for the translocation function (see Chap. 12). In the case of host recognition fibers, it is possible that the mismatch facilitates flexibility to scan for and attach to the viral receptor, as well as fiber removal upon binding, a step required to initiate the cascade of signals in both cell and virion for appropriate entry and/or genome delivery.

11.3.3 Special Vertices

Special (or singular) vertices have been found in many icosahedral dsDNA viruses. Singular vertices play key roles in genome packaging and ejection; they may also represent initial or final points in the assembly pathway of the shell. They represent a rupture of icosahedral symmetry (one vertex different from the other 11), and often include a symmetry mismatch (protein with non-fivefold symmetry occupying a fivefold coordinated position in the icosahedral net). Dodecameric proteins involved in genome packaging are found in a single vertex in tailed bacteriophages (see Sect. 11.3.4, and Chap. 12), as well as in herpesviruses [13, 14]. The best characterized case of special vertex is the portal in tailed bacteriophages, which connects the icosahedral head with the conspicuous tail that is characteristic of this viral family.

Giant dsDNA viruses infecting eukaryotic microorganisms also have singular vertices. *Paramecium bursaria Chlorella* virus-1 (PBCV-1) has a 190 nm diameter icosahedral capsid surrounding a lipid bilayer and dsDNA genome. A 250 Å long spike protrudes from one of the capsid vertices [8] (Figs. 11.4c, d). The peripentonal capsomers around the singular vertex seem to be structurally different from the rest. A ring-shaped density is observed near the singular vertex inside the capsid, which may correspond to a portal structure involved in genome packaging; however, there is no indication of symmetry mismatch between the fivefold vertex and this ring [29]. The spike is too thin to be used as a DNA ejection tube; besides PBCV-1 is thought to deliver its genome into the host by fusion of the internal membrane with the host one. It has been proposed that the function of the PBCV-1 spike is to puncture the cell wall to initiate the fusion process. The capsid side holding the spike is disassembled upon attachment to the host [29].

The giant Mimivirus has a 500 nm large icosahedral capsid structurally related to those of adenovirus, bacteriophage PRD1, and PBCV-1, covered by 125 nm long fibers. Early images of Mimivirus showed a starfish-shape feature with five arms reaching from one of the vertices to the five neighbouring ones (Fig. 11.4e). The arms of the starfish are inserted between adjacent facets, opening a gap between them. The starfish is an independent macromolecular assembly that remains together when detached from the virion [21], and is the only part of the capsid

not covered by fibers. When Mimivirus enters the cell by phagocytosis, a remarkable structural change occurs, whereby the five icosahedral facets in contact with the starfish feature open, leading to the structure called "stargate" [35]. The internal viral membrane is extruded through the stargate, to fuse with the phagosome membrane and release the viral DNA into the cytosol (Fig. 11.4f). Tailed phages use their special vertex both for genome packaging and delivery (Sect. 11.3.4; Chaps. 12, 17). In Mimivirus however, the stargate vertex is used for genome delivery, but not for packaging, which occurs instead *via* an aperture located in the icosahedral facet.

Remarkably, the asymmetry originated by the singular vertex in both PBCV-1 and Mimivirus reflects in an asymmetry of the internal virion contents [8, 21]. The viral genome and surrounding membrane do not occupy the full internal volume of the capsid. Rather, a gap exists between the DNA core and the side of the capsid containing the special vertex. This gap forms a pocket where viral enzymes required for cell membrane penetration may be contained. It may also contain structural elements required to precisely determine the asymmetric location and shape of the genome within the virion. However, these elements have not been identified or imaged so far.

Other viruses, such as PRD1 or adenovirus, have been reported to have singular vertices, based on genetic, biochemical and immunolabeling assays [40, 41]. However, for these viruses no structural information on the singular vertex is available yet, possibly due to the lack of large conspicuous features (such as tails) that would help calculation of three-dimensional (3D) maps without imposing full icosahedral symmetry.

11.3.4 The Extreme Case: Heads, Tails and Baseplates (Tailed Phages)

Tailed bacteriophages (order *Caudovirales*) are among the best described and more complex of the non-enveloped viruses. Their virions are composed by several functionally specialized morphological units, arranged according to different symmetries and connected *via* multiple symmetry mismatches (see also Chap. 17).

Bacteriophage capsids (heads) contain the dsDNA viral genome. They are icosahedrally ordered, but details vary among the different viruses. For example, HK97 or T7 phages have a strictly icosahedral head with a $T = 7$ net and a single major capsid protein occupying both the sixfold and fivefold coordinated positions. Others, like T4, have elongated, prolate icosahedral heads; two different proteins form the hexameric and pentameric capsomers, and several minor proteins are located at specific positions on the head surface [42] (Fig. 11.5a). One of the 12 vertices in the head is different from the other 11. Instead of a regular penton, it contains the portal structure or connector. This specialized structure is critical during assembly, since it contains the machinery needed to select the viral genome

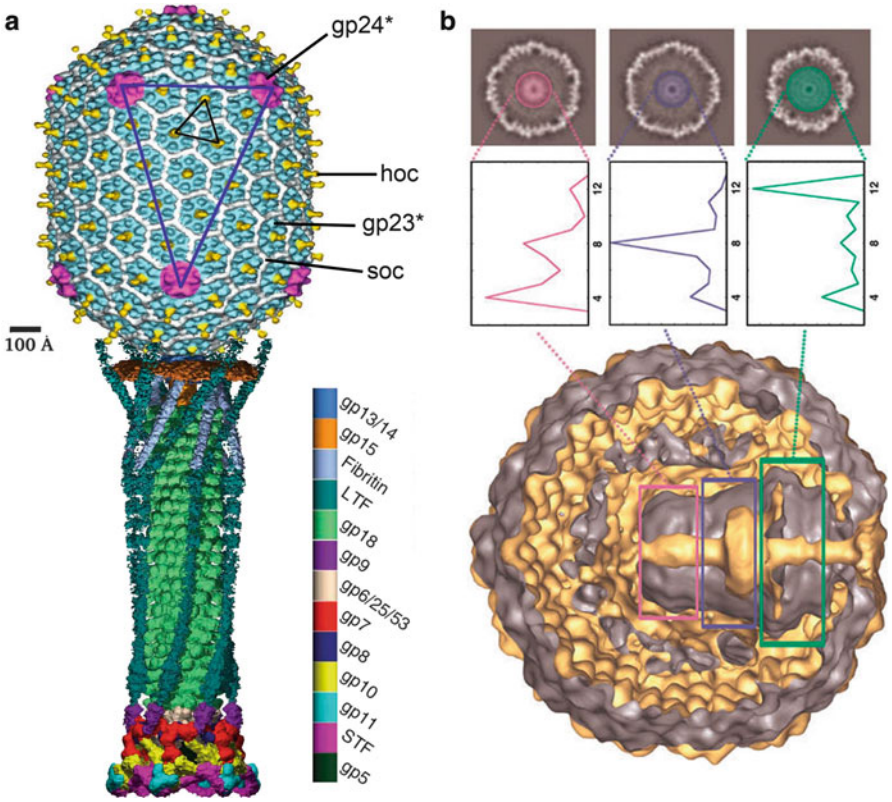


Fig. 11.5 Structure of tailed bacteriophages. (**a**) Bacteriophage T4 virion. The structures of the head and tail/baseplate have been solved separately, and merged to compose a representation of the complete virion. The tail is shown in its extended conformation (Modified from [43, 44]. With permission). (**b**) 3D map of the bacteriophage T7 procapsid, showing the internal core complex. The plots show the dominant symmetry for each core region (Reproduced from [45]. With permission). Diameter of the T7 procapsid = 510 Å

and pump it into the head (see Chap. 12). In addition, it links the head to the feature that differentiates these viral families: the tail.

Bacteriophages in the *Caudovirales* order are further classified in three groups, depending on the tail morphology: long, contractile (*Myoviridae*); long, non-contractile (*Siphoviridae*); and short, non-contractile (*Podoviridae*). In *Myoviridae* (representative: T4), the tail is composed by two layers of protein, one of them (the outer one) contractile [46] (Fig. 11.5a). The tail connects the head to a distal structure called the baseplate, formed by at least 16 different proteins in T4. Fibers with different lengths protrude from the baseplate; fibers may also be present at the portal region and the icosahedral head. In siphoviruses (representative: phage λ), the tail lacks the outer contractile sheath. Research on the structure of non-contractile tails has unveiled a crucial structural element: the tape measure protein,

whose length determines that of the tail by limiting the stacking of tail protein rings to a defined number. The baseplate composition is variable among the different viruses in this family, with some having only a simpler element called the tail tip complex. Fibers project both laterally from the periphery or longitudinally from the very the tip of the tail tip complex [47]. Podoviruses (representatives: T7, P22) have short tails, with a fiber complement that may include long, thin fibers (T7) or thick spikes (P22) [48]. Fibers, tails and baseplates or tail tips form the complex machinery required to initiate infection by recognizing and attaching to the host, and delivering the viral genome through the many layers protecting the bacterial cell (see Chap. 17).

Tailed bacteriophages deviate from the icosahedral symmetry due to their conspicuous genome delivery apparatus; additionally, they are a compendium of symmetry mismatches. Icosahedral (prolate or not) heads have a singular vertex where a fivefold symmetric capsomer is replaced by a 12-fold ring of the portal protein [13]. The portal complex is connected to the tail, which in general follows sixfold symmetry along the tube and baseplate. In the case of myoviruses however, a further mismatch may exist, since the contractile sheath presents helical symmetry, and it is not yet clear if the inner tube follows the arrangement of the sheath or the sixfold symmetry observed in non-contractile tails [46]. Additionally, some podoviruses such as T7, incorporate an internal proteic structure referred to as the core. This structure grows from the portal vertex towards the capsid center, and is thought to serve as a spindle for wrapping the DNA. In T7, the core presents eightfold and fourfold symmetries [45] (Fig. 11.5b). Finally, another symmetry mismatch may appear when the packaging motor binds to the portal vertex in the prohead during encapsidation. Reported oligomeric states for components of packaging motors include pentamers (T4 gp17, Φ29 pRNA), octamers (SF6 small terminase) and tetramers (λ terminase). However, for some of these motors it is not clear if the oligomerization states found in recombinant proteins are the same than in the prohead (immature capsid) context [13].

11.4 Asymmetric Virus Particles

11.4.1 Brick-Shaped Viruses

Poxviruses are large, enveloped dsDNA viruses apparently lacking any kind of high order symmetry in their capsids. The representative of the poxvirus family is vaccinia virus (Fig. 11.6a). The complex composition, large size, and asymmetric organization; plus the sensitivity of these viruses to the different preservation methodologies for EM, have caused the architecture of vaccinia to be a subject of debate for a long time. To complicate things even more, vaccinia virions exist in three different infectious forms carrying a different number of envelopes: mature virions, wrapped virions, and extracellular virions. The mature virion of vaccinia is

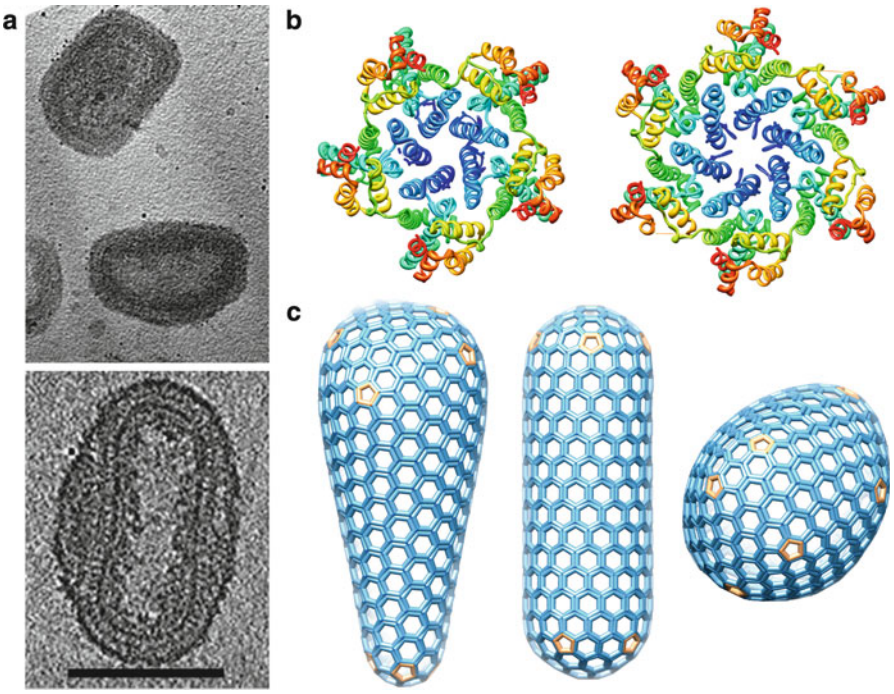


Fig. 11.6 Asymmetric and pleomorphic viruses. (**a**) Cryo-electron tomography reconstruction of mature vaccinia virions. *Top*: central section showing two virions in different orientations. *Bottom*: Central section of a virion where the dumbbell shape of the core can be appreciated. The bar represents 2,000 Å (Reproduced from [49]. With permission). (**b**) Structure of HIV CA protein assembled as a pentamer (*left*; PDB ID 3P05) or a hexamer (*right*; PDB ID 3H4E). (**c**) Example of fullerene-like objects generated from 12 pentamers and a variable number of hexamers

a brick-shaped enveloped particle with approximate dimensions 350 × 250 × 150 nm (varying depending on the imaging technique used), composed by at least 75 different proteins [17]. Mature virions are intracellular forms of the viral particle. Wrapped virions are also found in the cells, and consist of mature virions surrounded by two additional membranes derived from the Golgi cisternae. Wrapped virions leave the cell by fusing with the cell membrane, leaving one of their envelopes behind, to produce the extracellular virion.

The 200 kbp genome of vaccinia is contained in a core with an elongated dumbbell shape, surrounded by a protein capsule (core wall). The core also contains a variety of viral enzymes involved in RNA metabolism, required for the virus to replicate in the cytosol. The outer part of the core wall has striated appearance (palisade layer) while the inner part is smooth. It is not known if these different appearances are due to the existence of two chemically different layers or if there is only one asymmetrically organized layer. The extremes of the dumbbell rest adjacent to the envelope, while the central part is surrounded by electron-dense material (lateral bodies) of unknown function. *In vitro* disruption studies suggest

that the dsDNA in the core is in complex with condensing proteins [50]. However, the condensing proteins have not been identified yet.

11.4.2 Pleomorphic Viruses

Pleomorphic viruses not only do not follow high symmetry rules when forming the infectious particle, but may even adopt a wide range of sizes, shapes and composition from particle to particle, making each virion unique. Because of their intrinsic variability, the structural organization of pleomorphic viruses cannot be deduced from structural biology techniques based on averaging data from many identical virions, such as X-ray crystallography or cryo-EM analyses. The advent of electron tomography to visualize single virus particles (see Chap. 3) has started to reveal the architectural details of this kind of macromolecular machines, which includes many important pathogens for humans. Examples of pleomorphic viruses include retroviruses (HIV); orthomyxoviruses (influenza); coronaviruses (SARS-coronavirus); and paramyxoviruses (measles) [2, 51–53]. In addition, atomic force microscopy can be used for surface visualization of any kind of single virus particles (see Chap. 8) and holds great potential for imaging pleomorphic viruses in liquid, in close to physiological conditions. Possible deformations by adhesion to a solid base should be, however, considered in this case.

Pleomorphism is most pronounced among enveloped viruses, since the lipid envelope readily adapts different shapes and sizes. But also proteins with a tendency to form symmetric aggregations can give rise to pleomorphic capsids. The capsid protein of retroviruses (CA) can assemble into either hexamers or pentamers, in much the same way as capsid proteins of icosahedral viruses (Fig. 11.6b). Recombinant CA forms only hexamers in certain conditions, giving rise to tubular oligomers or flat, ordered sheets; while when pentameric oligomerization is enforced, $T = 1$ icosahedral particles are formed [54, 55]. However, when CA hexamers and pentamers associate to form the closed mature capsid that contains the nucleocapsid complex including the ssRNA genome, they do it in such manner that the pentamers are not distributed regularly within the hexamer lattice. Even if a fixed number of 12 pentamers is incorporated into each capsid, the asymmetry of their distribution results in asymmetrical structures that can adopt shapes ranging from roughly spherical to roughly conical, and can be modeled using the geometrical principles governing fullerene cones (Fig. 11.6c).

11.4.3 A Glimpse of the Weird Shapes of Archaeal Viruses

In the last years numerous new microorganisms living in extreme environments have been described, and with them their corresponding infecting viruses [56]. The most abundant repertoire of archaeal viruses reported so far is that of dsDNA

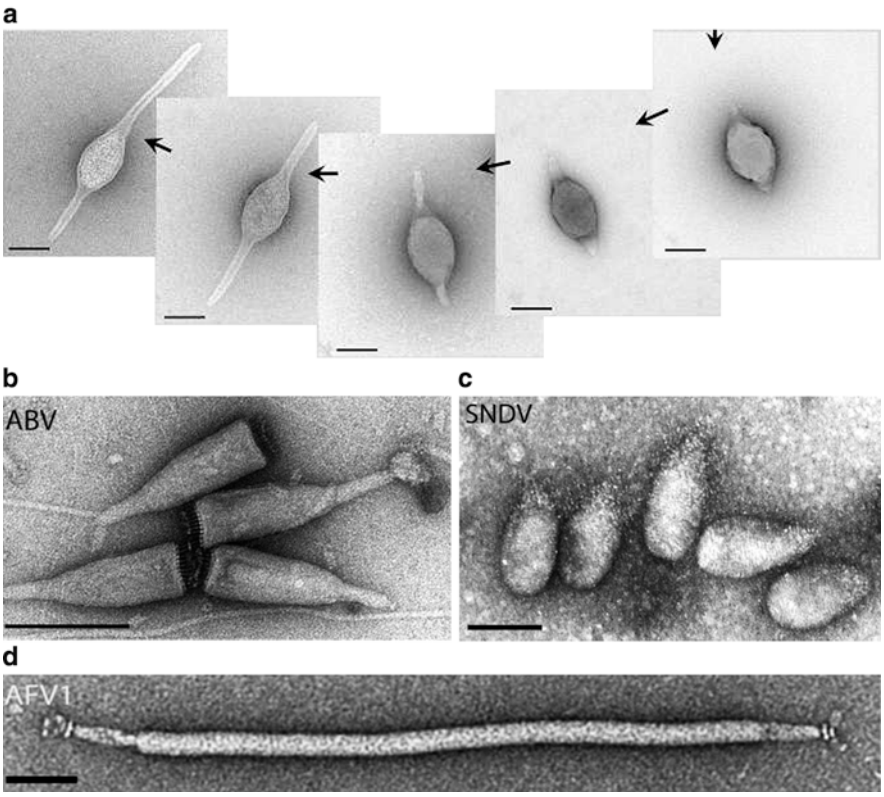


Fig. 11.7 Examples of archeoviruses with unusual shapes. (**a**) From *right* to *left*: extracellular elongation of the two tails in the fusiform ATV virion. (**b**) Bottle-shaped and (**c**) droplet-shaped viruses. (**d**) A Lipothrixvirus representative (AFV-1) with claw-shaped structures at both ends (Panel (**a**) reproduced from [60]; (**b**), (**c**) and (**d**) reproduced from [56]. With permission). Scale bars represent 100 nm

viruses. Only a few ssDNA archaeal viruses have been described, enclosing their genome in pleomorphic enveloped particles [57]. It is still not clear if RNA archaeal viruses exist [58]. Some dsDNA viruses infecting archaea follow the general architectural types previously known for bacterial and eukaryotic viruses. For example, *Sulfolobus* turreted icosahedral virus (STIV), *Haloharcula hispanica* SH1, or *Salisaeta* icosahedral phage 1 (SSIP-1) [59] are icosahedral, tailless viruses with an internal membrane, structurally similar to bacteriophage PRD1 [11] or the algae virus PBCV-1 [8]. Some archaeal viruses with tailed bacteriophage morphologies have also been reported [56].

Intriguingly, other viral families discovered in archaea are unique to this branch of life, with morphologies never observed before (Fig. 11.7). According to their overall organization, they are classified into fusiform, droplet or bottle-shaped, and linear viruses. Within each class, viruses with different genome types, sizes, and

bearing no sequence similarity can be found. Fusiform viruses are very abundant in habitats dominated by archaeal microorganisms. They have spindle-shaped virions with tails of variable length protruding from the spindle poles. One fusiform virus, *Acidianus* two-tailed virus (ATV) not only has an unusual shape, but is able to assemble new structural features after leaving the host cell. When propagated at temperatures slightly suboptimal for its host (75 °C), isolated virions look like ~0.2 μm long lemons. However, when temperature is raised in the absence of the host cell, these viruses grow two filamentous tails of variable lengths, one from each pole. This is the only known example of a virus with extracellular assembly, but it is likely that others exist [60]. The tails end in an anchor-like structure, thought to be involved in attachment to the host.

Acidianus bottle-shaped virus (ABV) and *Sulfolobus neozelandicus* droplet-shaped virus (SNDV) are the only known members of the two viral families termed *Ampullaviridae* and *Guttaviridae* [56]. The enveloped ABV virion contains a conical core formed by a supercoiled nucleoprotein filament. A brush of short filaments protrudes from the bottom of the bottle, but host attachment seems to occur at the opposite side of the virion. Little is known about the architecture of the SNDV virion, except for its droplet shape and the presence of a tuft of long fibers at its narrower pole. Finally, linear archaeal viruses can form stiff rods (*Rudiviridae*) or flexible filaments (*Lipothrixviridae*). Rudiviruses are relatively simple in composition, with no envelope and only a few proteins arranged in particles of variable length, usually related to that of the genome. Lipothrixviruses are enveloped, and the ends of their filamentous capsids are capped with structures of varied shapes (spider legs, pincers, bottle brushes), probably involved in attachment to the host.

11.5 Sophisticated Regulation of Assembly and Maturation

In simple viruses, assembly can occur either in a single step where the newly replicated nucleic acid associates with capsid protein subunits during co-assembly, or in a two-step process where an empty capsid is assembled first and the viral nucleic acid is packaged afterwards (see Chaps. 10, 12). For complex viruses, putting together the many different pieces in their proper places at the appropriate time requires elaborated regulation of the morphogenesis process. In the following sections we discuss some of the strategies used by viruses to achieve assembly of complex capsids. The additional steps required by enveloped viruses to coordinate assembly of proteic elements with recruitment of membranes from the cell will be described in Chap. 14.

11.5.1 Separate Assembly Lines

In viruses with complex chemical composition, the different morphological components are often built separately, forming subassemblies that will be later

put together along carefully regulated pathways. Some of these subassembly reactions may also require chaperones, either of cellular or viral origin, as is also the case for some simpler viruses (Chap. 10). For example, in adenovirus, capsid protein oligomers are formed in the cytosol before being transported to the nucleus, where viral assembly takes place. This preassembly step includes hexon trimerization, which requires a viral chaperone (L4 100K) [61]; and piecing together the vertex complex, composed by a pentamer of penton base bound to a trimer of the fiber protein [4]. In adenovirus, however, the precise temporal order of incorporation of major and minor capsid proteins is not yet understood. Similarly, in herpesvirus hexamers and pentamers of the major capsid protein VP25 are formed previous to particle assembly.

The best described examples of subassembly formation and integration into a virion come from the order *Caudovirales*, more specifically from the *Myoviridae* prototype bacteriophage T4 [42]. In these long tailed bacteriophages, the head, fibers, and tail form separately (Fig. 11.8). Bacteriophage T4 head assembly starts from an initiation complex containing the portal protein gp20 bound to a cellular membrane. This complex recruits the components of a scaffolding core composed by eight different types of viral proteins, among them the main scaffolding protein gp22 and the viral protease (gp21). This core will later be coated by the hexameric and pentameric capsomers (gp23 and gp24) to form the procapsid, or prohead. Once coating is complete, gp21 is activated and cleaves all components of the procapsid except the portal protein. The prohead is detached from the membrane and most of the cleaved peptides exit the particle, opening the way for entry of the viral genome. Genome packaging leads to the large structural rearrangements involved in capsid maturation (Sect. 11.5.3 and Chap. 13), to produce the head in its final form. Finally, the head completion (or neck) proteins gp13, gp14, gp2 and gp4 attach to the portal to form the interface between the head and the tail.

The tail is in turn formed from several preassembled pieces. For contractile tails, the baseplate is assembled first, and used as a seeding point for assembly of the inner tube and contractile sheath. In bacteriophage T4, tail assembly involves 19 different proteins and seven viral chaperones. To form the T4 baseplate, proteins gp6, gp7, gp8, gp10, gp11, gp25 and gp53 assemble in the form of hetero-oligomeric wedges. Six wedges bind around a central hub containing gp5 and gp27. Proteins gp9 and gp12 (the short tail fiber) are then inserted at the gaps between wedges, and the interface between wedges and hub is sealed by proteins gp48 and gp54. This seal is the starting point from which the gp19 inner tail tube will grow. The length of the tube is controlled by a tape measure protein gp29, which extends from the hub to the tube end where the tail capping protein gp3 will bind. The tail sheath gp18 assembles around the inner tube, and finally the tail terminator protein gp15 binds to gp3 and the last row of gp18 subunits, making the tail ready to bind to the neck proteins in the head.

The final stage of tail assembly is incorporation of the long tail fibers to the base plate. The fibers also assemble independently, even starting from separate

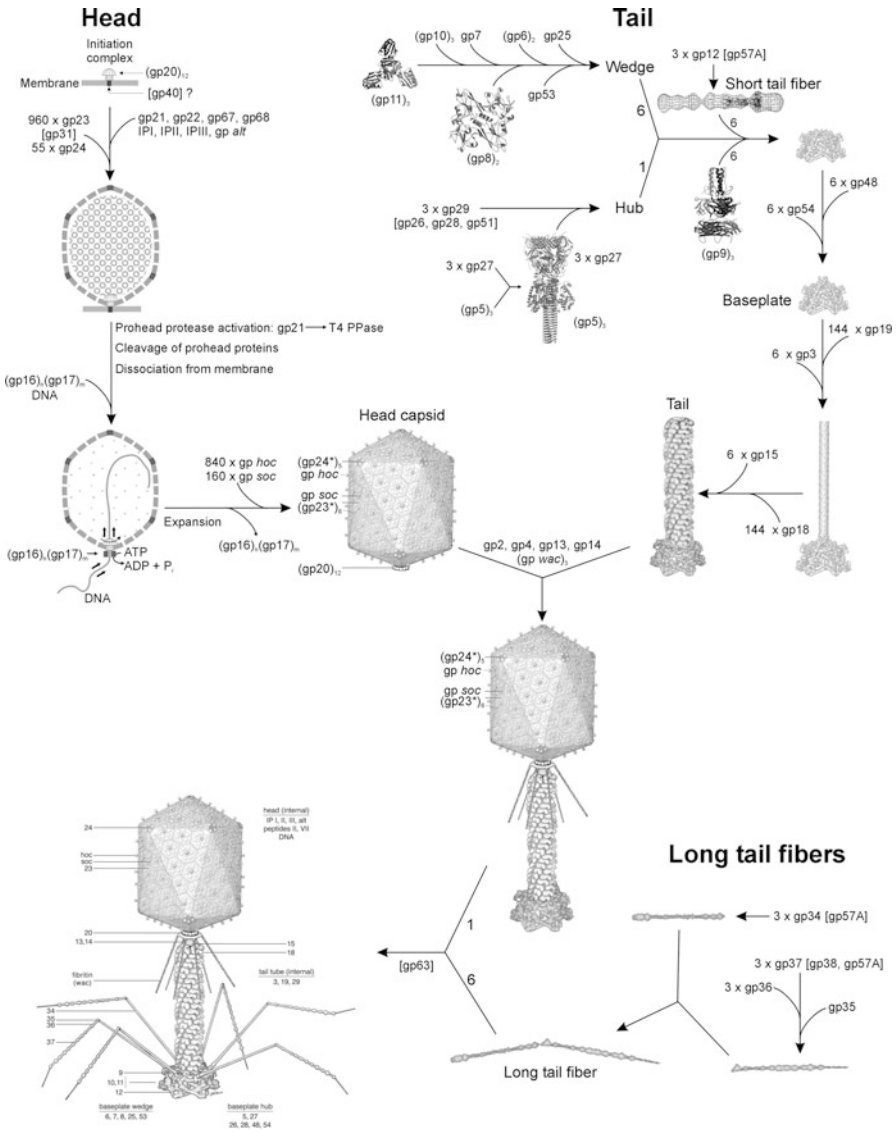


Fig. 11.8 Separate assembly lines. Schematic depiction of the complex bacteriophage T4 morphogenesis pathway (Reproduced from [42]. With permission)

subassemblies. The long tail fiber of bacteriophage T4 is kinked; the proximal part of the fiber before the kink is formed by a single protein (gp34), while the distal part contains three different proteins: gp35, gp36, and gp37. The proximal and distal parts assemble separately, and then join before attaching to the baseplate. All long

tail fiber proteins form trimers, except monomeric gp35 (notice the symmetry mismatch) that sits at the interface between the distal and proximal half fibers. Interestingly, the long fibers cannot join the tail until it is bound to the DNA-filled head. A similar assembly pathway has been described for non-contractile long tails. In podoviruses however, the short tail is not assembled as a separate entity, but it grows outward from the portal vertex on the virion capsid. The careful temporal regulation of these assembly processes is proved by the fact that, when any of the structural proteins is absent, viral morphogenesis is interrupted and the assembly intermediates previous to the disrupted step accumulate in the cell.

11.5.2 Scaffolding

Scaffolding elements are crucial for accurate assembly of large viral capsids. They are present in assembly intermediates (*e.g.* procapsids), but absent in the final, infective product. Their role is to facilitate interactions between capsid elements at the early stages of assembly, by promoting nucleation – that is, putting together the viral proteins that may be highly diluted within the crowded cellular context. Scaffolds are also thought to stabilize weak interactions at the initial stages of assembly [62], while simultaneously allowing flexibility for mistake corrections. This last function is most important in large capsids, where the number of interactions to be checked for errors is correspondingly large. Finally, scaffolds have a role in determining the size and shape of viral capsids.

The most studied scaffold proteins are those present in tailed bacteriophages. For example, bacteriophage P22 (*Podoviridae*) scaffold is a 33 kDa protein. In the early stages of P22 morphogenesis, a procapsid is formed by 415 copies of the capsid protein, with approximately 300 molecules of scaffold inside. Unlike the capsid protein, the scaffold does not follow icosahedral symmetry; therefore, little is known about its organization in the assembly intermediate [45]. Scaffold proteins have been quite refractory to structural studies. Nuclear magnetic resonance and crystallographic studies on the scaffold of P22 and Φ29 indicate that they have a helical fold. Biophysical analyses indicate that many of them share an extended, rod-like shape and a tendency to dimerize in solution [63]. However, an equilibrium between different oligomeric forms seems to be required to achieve correct capsid assembly. Kinetic studies have revealed that in phage P22, scaffold is predominantly a dimer during assembly, but the presence of free monomers is absolutely required to complete the head. Kinetically trapped intermediates are observed when monomers are depleted by decreasing the ionic strength, while restoring it eliminates the trap and allows elongation to proceed. Phage scaffolds are usually ejected from the procapsid immediately before genome packaging. In P22 and Φ29, the intact protein exits the shell, and can be recycled in a new round of assembly. In other cases, the scaffold is removed *via* cleavage by a viral-encoded protease.

In spite of their apparently simple organization, some small bacteriophages, such as the Microvirus representative ΦX174 (T = 1), encode both internal and external scaffold proteins [63]. The ΦX174 internal scaffold protein (protein B) helps in the early stages of assembly by preventing aggregation of the capsid protein F into aberrant oligomers, and ensuring the recruitment of the vertex spike protein G. The C-terminal region of protein B (24 aminoacids) interacts with the capsid and can be observed in the crystal structure of the procapsid, while the rest is disordered and appears to be largely tolerant to mutations. On the contrary, the external scaffold protein D is highly ordered and sensitive to mutations. Protein D is absolutely required for elongation (to assemble capsid pentamers into a spherical particle), while protein B helps to make assembly efficient but is not strictly required: in the absence of B, viral particles can be formed, but the process requires overexpression of protein D and takes as much as ten times longer than in the presence of both scaffolds. It is believed that scaffold redundancy confers an evolutionary advantage to ΦX174 by facilitating extremely rapid replication cycles.

In the absence of scaffold, many phage capsid proteins have been observed to self-assemble into aberrant oligomers (tubes, elongated shells, T = 1 icosahedrons). It follows from these observations that scaffolding proteins are involved in determining the correct curvature in the interactions between capsomers, so that they can form a closed shell of the appropriate size to hold the viral genome. A remarkable proof of this size determination role comes from the P2/P4 phage system [64]. Phage P2 is a member of the *Myoviridae* family, with an assembly pathway similar to P22 that includes a T = 7 procapsid formed with the help of an internal scaffolding protein, gpO. Remarkably, P2 gpO combines the function of scaffold and protease, able to cleave itself and other shell components at later maturation stages. In the presence of P4, however, the whole assembly pathway is altered. P4 is a replicon that can exist either as a prophage or a free plasmid, and does not code for any major structural proteins. However, P4 is able to sequester players of the P2 assembly line to form small T = 4 capsids where only its smaller genome can fit, excluding that of P2 (Fig. 11.9). This parasitic process is achieved by synthesis of a P4-encoded scaffold protein, Sid (SIze Determination protein). Sid forms an external scaffold on P2 assembly intermediates, forcing a narrower curvature and therefore a smaller icosahedral net. Unlike the internal scaffolds, P4 Sid forms a dodecahedral ordered cage that can be observed in cryo-electron microscopy reconstructions.

Although one can generally speak about scaffolding proteins, and indeed many viruses have such proteins, scaffolding functions can also be performed by flexible regions of the capsid proteins, which establish interactions during assembly that are later removed *via* conformational changes or cleavage by viral proteases. For example, bacteriophage HK97 (*Siphoviridae*) does not encode a scaffold protein. Instead, a 103 residue stretch at the N-terminus of the capsid protein, known as the delta-domain, performs the scaffold function [65]. The delta domain is located towards the interior of the capsid and mediates interactions between capsomers during assembly. Once the procapsid is completed, and before the DNA is packaged, the delta domain is cleaved out by the viral protease, allowing the transition to

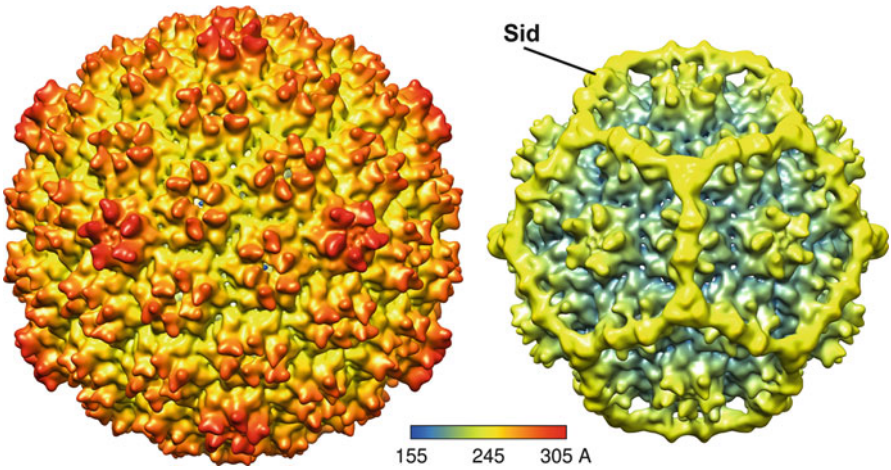


Fig. 11.9 Capsid size determination by scaffolding proteins: the P2/P4 phage system. *Left*: bacteriophage P2 T = 7 procapsid (EMDB ID EMD_5406). *Right*: the parasitic P4 T = 4 procapsid (EMDB ID EMD_5405). Note the Sid external scaffold restricting the shell size. The view is along a twofold icosahedral axis. Original maps have been filtered to a lower resolution for clarity. Color key indicates color changes with map radius

the next assembly intermediate [62]. Adenovirus may use a dual system, including both a separate scaffolding protein [66] and flexible regions of minor capsid proteins removed by the viral protease during maturation [67].

11.5.3 Maturation

In complex viruses, piecing together a number of proteins into a capsule to host the genome is far from producing the final, infectious form. Instead, the newly formed particles (procapsids) need to undergo a series of morphological and/or stability changes to acquire their full infectious potential. This process is known as maturation. There is a double goal for maturation on the viral cycle: first, to produce virions stable enough to protect the genome from aggressive conditions in the extracellular milieu; and second, to prepare the viral particle for correct delivery of the genome into the new host cell.

In dsDNA bacteriophages, maturation encompasses large structural changes and protein rearrangements in the capsid, concomitant with scaffold removal and genome packaging. The capsid changes from a weak, labile object to a highly stable shell, ready to withstand the high internal pressure imposed by the tightly packed DNA inside [62] (see Chaps. 9, 12, 18, and 19). In other viruses, such as polio [68] or adenovirus [67], maturation does not end with an extremely stable shell, but with a metastable one. This difference with respect to the bacteriophage case is likely related to the mode of infection of these eukaryotic viruses. Instead of

ejecting the genome across the plasma membrane leaving the capsid behind, polio and adenovirus are internalized in the cell, and must be disassembled within in a concerted fashion to ensure exposure of the genome at the appropriate place and time for successful replication. Maturation prepares these viruses to start the programmed uncoating sequence upon reception of the appropriate signal, for example attachment to the receptor, or pH changes along the endocytosis pathway. Interestingly, in adenovirus maturation is related to genome packaging in quite a unique way. The adenoviral protease, which is the main maturation agent, is packaged together with the viral genome thanks to its dsDNA binding ability, and uses the dsDNA itself as a cofactor to increase its catalytic activity several orders of magnitude [4].

Maturation processes are not restricted to icosahedral capsids: pleomorphic enveloped viruses such as retroviruses also undergo extensive structural rearrangements to become fully infectious [2]. A more extensive discussion on maturation for several different viruses can be found in Chap. 13.

11.6 Perspectives and Conclusions

In the past decade, structural studies on complex viruses have greatly benefited from technical improvements in structural biology techniques such as those described in Chaps. 3, 4, 5, 6, 7, 8, and 9. As more details are known, it is becoming clear that many complex viruses that infect hosts far apart in evolution share common structural solutions.

For example, adenovirus, which infects vertebrates, has a striking structural similarity to PRD1, a bacteriophage with an internal membrane. The parallels between adenovirus and PRD1 extend from their DNA replication mechanisms, to their capsid architecture and the folding of their major capsid proteins [3, 11]. In the last years, more members of the PRD1-adenovirus family have been described or predicted, and the lineage now extends from viruses infecting bacteria or archaea, to the large nucleo-cytoplasmic DNA viruses such as Asfarvirus, Iridovirus and the giant Mimivirus [9]. All these viruses are built from the same kind of double 8-stranded β-barrel, pseudo-hexagonal capsomers arranged in different tiling systems, with triangulation numbers ranging between $T = 21$ and $T = 169$, and reaching up to $972 < T < 1{,}200$ for the giant Mimivirus [21]. Intriguingly, even a scaffold protein of the non-icosahedral vaccinia virus folds as a double barrel pseudo-hexamer, indicating a possible common ancestor with icosahedral dsDNA viruses [69].

Adenovirus and PRD1 are not the only cases indicating an evolutionary relationship between animal and bacterial viruses. Herpesviruses, which infect all sorts of animal organisms, share many structural characteristics with tailed bacteriophage [14]. They follow a similar assembly pathway, starting from an empty procapsid formed with the help of scaffold, and maturing to a more angular shell *via* large structural rearrangements upon DNA packaging. Like tailed bacteriophage, one of

the vertices is different from the rest and contains a 12-fold symmetric portal structure involved in genome encapsidation. Herpesvirus capsid protein folds with a topology very similar to that of the HK97 phage family; and like tailed phages, the packed genome forms concentric shells when observed in icosahedrally averaged EM reconstructions. Finally, structural parallels also exist between Cystoviruses (dsRNA bacteriophage, representative Φ6) and Reoviruses [70].

The fact that many complex viruses with different hosts share a common structural solution has evolutionary implications. First, since the design has been conserved throughout time, even after all traces of sequence similarity have disappeared, it must be a highly efficient arrangement. Second, it suggests that the architecture was established in the early stages of evolution, before the branches of the evolutionary tree diverged into the three kingdoms known today (archaea, bacteria and eukarya). One could wonder, then, if all complex viruses existing today fall within a limited number of structural solutions selected by their success early in evolution [71]. However, discovery of the unique architectures of archaeal viruses indicates that other structural solutions exist. Advances in high throughput virus isolation and structural characterization techniques will contribute to clarify this question.

In summary, complex viruses incorporate a wide range of molecules into their capsids, including specialized host interaction, genome packaging and cementing proteins; and in some cases membranes, either internal or external. Accommodation of the different components often requires deviations from high order symmetry, from mismatches to pleomorphism; and involves complex regulation of the assembly dynamics. Key elements in this regulation are: separate assembly lines, scaffold elements, and maturation processes. Host and virus evolution probably act hand in hand to optimize viral particle structure and morphogenesis.

Acknowledgements Work in the San Martín lab is funded by grants BFU2010-16382 and FIS2010-10552-E/FIS2011-16090-E from the Ministerio de Ciencia e Innovación of Spain. José R. Castón (CNB-CSIC) is acknowledged for careful reading of the manuscript.

References and Further Reading

1. Stubbs G, Kendall A (2012) Helical viruses. Adv Exp Med Biol 726:631–658
2. Goff SP (2007) Retroviridae: the retroviruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA (eds) Fields virology, vol II. Lippincott Williams & Wilkins, Philadelphia, pp 1999–2069
3. Liu H, Jin L, Koh SB, Atanasov I, Schein S, Wu L, Zhou ZH (2010) Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science 329:1038–1043
4. San Martín C (2012) Latest insights on adenovirus structure and assembly. Viruses 4:847–877
5. Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. Mol Cell 17:121–135
6. Rux JJ, Kuser PR, Burnett RM (2003) Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J Virol 77:9553–9566

7. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
8. Cherrier MV, Kostyuchenko VA, Xiao C, Bowman VD, Battisti AJ, Yan X, Chipman PR, Baker TS, Van Etten JL, Rossmann MG (2009) An icosahedral algal virus has a complex unique vertex decorated by a spike. Proc Natl Acad Sci U S A 106:11085–11089
9. Krupovic M, Bamford DH (2008) Virus evolution: how far does the double beta-barrel viral lineage extend? Nat Rev Microbiol 6:941–948
10. Colby WW, Shenk T (1981) Adenovirus type 5 virions can be assembled *in vivo* in the absence of detectable polypeptide IX. J Virol 39:977–980
11. Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martín C, Burnett RM, Stuart DI, Bamford DH, Bamford JK (2004) Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432:68–74
12. Roizman B, Knipe DM, Whitley RJ (2007) Herpes simplex viruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA (eds) Fields virology, vol II. Lippincott Williams & Wilkins, Philadelphia, pp 2501–2601
13. Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. Adv Exp Med Biol 726:489–509
14. Cardone G, Heymann JB, Cheng N, Trus BL, Steven AC (2012) Procapsid assembly, maturation, nuclear exit: dynamic steps in the production of infectious herpesvirions. Adv Exp Med Biol 726:423–439
15. Huiskonen JT, Jaalinoja HT, Briggs JA, Fuller SD, Butcher SJ (2007) Structure of a hexameric RNA packaging motor in a viral polymerase complex. J Struct Biol 158:156–164
16. Trask SD, McDonald SM, Patton JT (2012) Structural insights into the coupling of virion assembly and rotavirus replication. Nat Rev Microbiol 10:165–177
17. Condit RC, Moussatche N, Traktman P (2006) In a nutshell: structure and assembly of the vaccinia virion. Adv Virus Res 66:31–124
18. Germond JE, Hirt B, Oudet P, Gross-Bellark M, Chambon P (1975) Folding of the DNA double helix in chromatin-like structures from simian virus 40. Proc Natl Acad Sci U S A 72:1843–1847
19. Tweeten KA, Bulla LA, Consigli RA (1980) Characterization of an extremely basic protein derived from granulosis virus nucleocapsids. J Virol 33:866–876
20. Giberson AN, Davidson AR, Parks RJ (2012) Chromatin structure of adenovirus DNA throughout infection. Nucleic Acids Res 40:2369–2376
21. Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG (2009) Structural studies of the giant mimivirus. PLoS Biol 7:e92
22. Ruigrok RW, Crepin T, Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. Curr Opin Microbiol 14:504–510
23. Palese P, Shaw ML (2007) Orthomyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA (eds) Fields virology, vol II. Lippincott Williams & Wilkins, Philadelphia, pp 1647–1689
24. Huiskonen JT, Butcher SJ (2007) Membrane-containing viruses with icosahedrally symmetric capsids. Curr Opin Struct Biol 17:229–236
25. Grunewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, Steven AC (2003) Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science 302:1396–1398
26. Forster F, Medalia O, Zauberman N, Baumeister W, Fass D (2005) Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. Proc Natl Acad Sci U S A 102:4729–4734
27. Butcher SJ, Manole V, Karhu NJ (2012) Lipid-containing viruses: bacteriophage PRD1 assembly. Adv Exp Med Biol 726:365–377

28. Grahn AM, Daugelavicius R, Bamford DH (2002) Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. Mol Microbiol 46:1199–1209
29. Zhang X, Xiang Y, Dunigan DD, Klose T, Chipman PR, Van Etten JL, Rossmann MG (2011) Three-dimensional structure and function of the paramecium bursaria chlorella virus capsid. Proc Natl Acad Sci U S A 108:14837–14842
30. Tulman ER, Delhon GA, Ku BK, Rock DL (2009) African swine fever virus. Curr Top Microbiol Immunol 328:43–87
31. Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24
32. Baker M, Prasad BV (2010) Rotavirus cell entry. Curr Top Microbiol Immunol 343:121–148
33. Danthi P, Guglielmi KM, Kirchner E, Mainou B, Stehle T, Dermody TS (2010) From touchdown to transcription: the reovirus cell entry pathway. Curr Top Microbiol Immunol 343:91–119
34. Huiskonen J, Manole V, Butcher S (2007) Tale of two spikes in bacteriophage PRD1. Proc Natl Acad Sci U S A 104:6666–6671
35. Zauberman N, Mutsafi Y, Halevy D, Shimoni E, Klein E, Xiao C, Sun S, Minsky A (2008) Distinct DNA exit and packaging portals in the virus acanthamoeba polyphaga mimivirus. PLoS Biol 13:e114
36. van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature 401:935–938
37. Liu H, Wu L, Zhou ZH (2011) Model of the trimeric fiber and its interactions with the pentameric penton base of human adenovirus by cryo-electron microscopy. J Mol Biol 406:764–774
38. Chappell JD, Prota AE, Dermody TS, Stehle T (2002) Crystal structure of reovirus attachment protein σ1 reveals evolutionary relationship to adenovirus fiber. EMBO J 21:1–11
39. Hess M, Cuzange A, Ruigrok RWH, Chroboczek J, Jacrot B (1995) The avian adenovirus penton: two fibres and one base. J Mol Biol 252:379–385
40. Christensen JB, Byrd SA, Walker AK, Strahler JR, Andrews PC, Imperiale MJ (2008) Presence of the adenovirus IVa2 protein at a single vertex of the mature virion. J Virol 82:9086–9093
41. Gowen B, Bamford JK, Bamford DH, Fuller SD (2003) The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. J Virol 77:7863–7871
42. Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG (2003) Structure and morphogenesis of bacteriophage T4. Cell Mol Life Sci 60:2356–2370
43. Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov VV, Rossmann MG (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. Nat Struct Mol Biol 12:810–813
44. Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG (2004) Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci U S A 101:6003–6008
45. Agirrezabala X, Martin-Benito J, Caston JR, Miranda R, Valpuesta JM, Carrascosa JL (2005) Maturation of phage T7 involves structural modification of both shell and inner core components. EMBO J 24:3820–3829
46. Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. Adv Exp Med Biol 726:93–114
47. Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL (2012) Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol 726:115–142
48. Casjens SR, Molineux IJ (2012) Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol 726:143–179
49. Cyrklaff M, Risco C, Fernández JJ, Jiménez MV, Estéban M, Baumeister W, Carrascosa JL (2005) Cryo-electron tomography of vaccinia virus. Proc Natl Acad Sci U S A 102:2772–2777

50. Kuznetsov Y, Gershon PD, McPherson A (2008) Atomic force microscopy investigation of vaccinia virus structure. J Virol 82:7551–7566
51. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, Steven AC (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci U S A 103:19123–19127
52. Barcena M, Oostergetel GT, Bartelink W, Faas FG, Verkleij A, Rottier PJ, Koster AJ, Bosch BJ (2009) Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirion. Proc Natl Acad Sci U S A 106:582–587
53. Liljeroos L, Huiskonen JT, Ora A, Susi P, Butcher SJ (2011) Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions. Proc Natl Acad Sci U S A 108:18085–18090
54. Pornillos O, Ganser-Pornillos BK, Yeager M (2011) Atomic-level modelling of the HIV capsid. Nature 469:424–427
55. Cardone G, Purdy JG, Cheng N, Craven RC, Steven AC (2009) Visualization of a missing link in retrovirus capsid assembly. Nature 457:694–698
56. Pina M, Bize A, Forterre P, Prangishvili D (2011) The archeoviruses. FEMS Microbiol Rev 35:1035–1054
57. Pietila MK, Atanasova NS, Manole V, Liljeroos L, Butcher SJ, Oksanen HM, Bamford DH (2012) Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. J Virol 86:5067–5079
58. Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M (2012) Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated yellowstone hot springs. J Virol 86:5562–5573
59. Aalto AP, Bitto D, Ravantti JJ, Bamford DH, Huiskonen JT, Oksanen HM (2012) Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing salisaeta icosahedral phage 1. Proc Natl Acad Sci U S A 109:7079–7084
60. Haring M, Vestergaard G, Rachel R, Chen L, Garrett RA, Prangishvili D (2005) Virology: independent virus development outside a host. Nature 436:1101–1102
61. Hong SS, Szolajska E, Schoehn G, Franqueville L, Myhre S, Lindholm L, Ruigrok RW, Boulanger P, Chroboczek J (2005) The 100K-chaperone protein from adenovirus serotype 2 (subgroup C) assists in trimerization and nuclear localization of hexons from subgroups C and B adenoviruses. J Mol Biol 352:125–138
62. Johnson JE (2010) Virus particle maturation: insights into elegantly programmed nanomachines. Curr Opin Struct Biol 20:210–216
63. Prevelige PE, Fane BA (2012) Building the machines: scaffolding protein functions during bacteriophage morphogenesis. Adv Exp Med Biol 726:325–350
64. Marvik OJ, Sharma P, Dokland T, Lindqvist BH (1994) Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. Virology 200:702–714
65. Huang RK, Khayat R, Lee KK, Gertsman I, Duda RL, Hendrix RW, Johnson JE (2011) The prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation. J Mol Biol 408:541–554
66. Hasson TB, Ornelles DA, Shenk T (1992) Adenovirus L1 52- and 55-kDa proteins are present within assembling virions and colocalize with nuclear structures distinct from replication centers. J Virol 66:6133–6142
67. Pérez-Berná AJ, Ortega-Esteban A, Menéndez-Conejero R, Winkler DC, Menéndez M, Steven AC, Flint SJ, de Pablo PJ, San Martín C (2012) The role of capsid maturation on adenovirus priming for sequential uncoating. J Biol Chem 287:31582–31595
68. Hogle JM (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways. Annu Rev Microbiol 56:677–702
69. Bahar MW, Graham SC, Stuart DI, Grimes JM (2011) Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. Structure 19:1011–1020
70. Bamford DH, Burnett RM, Stuart DI (2002) Evolution of viral structure. Theor Popul Biol 61:461–470
71. Abrescia NG, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral universe. Annu Rev Biochem 81:795–822

Further Reading

Agbandje-McKenna M, McKenna R (2011) Structural virology. RSC Publishing, Cambridge
Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2009) Principles of virology. ASM Press, Washington, DC
Rixon FJ, Chiu W (2003) Studying large viruses. Adv Protein Chem 64:379–408
Rossmann MG, Rao VB (2012) Viral molecular machines. Adv Exp Med Biol, vol 726, Springer, New York

Also especially recommended for further reading are references [4, 9, 24, 32, 42, 54, 56, 62] *listed above*.

Chapter 12
Nucleic Acid Packaging in Viruses

Ana Cuervo, María I. Daudén, and José L. Carrascosa

Abstract Viruses protect their genetic information by enclosing the viral nucleic acid inside a protein shell (capsid), in a process known as genome packaging. Viruses follow essentially two main strategies to package their genome: Either they co-assemble their genetic material together with the capsid protein, or they assemble first an empty shell (procapsid) and then pump the genome inside the capsid with a molecular motor that uses the energy released by ATP hydrolysis. During packaging the viral nucleic acid is condensed to very high concentration by its careful arrangement in concentric layers inside the capsid. In this chapter we will first give an overview of the different strategies used for genome packaging to discuss later some specific virus models where the structures of the main proteins involved, and the biophysics underlying the packaging mechanism, have been well documented.

Keywords Bacteriophage • Capsid • Connector • DNA • Electron microscopy • Encapsidation • Ejection • Helical symmetry • Icosahedral symmetry • Maturation • Molecular motor • Nucleocapsid • Packaging • Portal • RNA • Shell • Terminase • X-ray diffraction • Virus

Abbreviations

ATP	Adenosine triphosphate
bp	Base pair
BPMV	Bean pod mottle virus
BTV	Bluetongue virus

A. Cuervo • M.I. Daudén • J.L. Carrascosa (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jlcarras@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_12,

Cdom	Carboxy domain
ds	Double-stranded
FHV	Flock house virus
HCMV	Human cytomegalovirus
mRNA	messenger RNA
Ndom	Amino domain
NPC	Nucleoprotein complex
nt	Nucleotides
NTP	Nucleotide triphosphate
PaV	Pariacoto virus
pRNA	prohead RNA
ss	Single stranded
STMV	Satellite tobacco mosaic virus
TMV	Tobacco mosaic virus

12.1 Introduction

Viruses are mainly composed by a carrier of genetic information (either DNA or RNA), and a protective and multifunctional container, usually made of proteins and, eventually, lipidic components [1] (see Chap. 2). Several main aspects have to be considered in virus construction. One is that it must consume the minimum possible genetic information, making use of geometric principles to build a large container using a limited set of proteins. In addition, the virus has to be easy to assemble to facilitate the highest possible progeny production but, at the same time, this assembly process has to incorporate mechanisms to select the viral genome and reject cellular components. Then, the virus particle must be able to actively participate in its release from the infected cell, the transfer to other possible hosts, their proper and accurate recognition and, finally, the delivery of their genetic information to start a new infection cycle.

It is clear that to carry out all these processes in so many different environments the virus particle must be a quite flexible vehicle able to trigger and undergo structural changes to provide at the appropriate moment the required functionality: the resistance needed to survive the harsh extracellular environment may be superfluous and even counter-productive in the well-controlled intracellular conditions. The propagation strategy is also reflected in the virus structure. In certain plants, the viral propagation allows each particle to carry partial genomic information, as very many viruses will co-infect the cells (*i.e.*, the genome is found multipartite in different viral particles). On the other hand, viruses infecting most animal cells present the whole genome either on a single molecule or in several segments within the same viral particle to ensure efficient infection. Also, the requirements to

penetrate the eukaryotic cell membrane (see Chaps. 15 and 16) are very different to those required to traverse the thick-walled bacterial envelope (see Chap. 17). Due to these and many other variables, different virus particles have evolved different survival strategies. Thus, viruses provide excellent model systems for the study of biotechnological solutions at the nanoscale.

Viruses offer a unique gradient of structural solutions to encapsidate and protect their genetic material, from the simplest ones to those revealing unexpected levels of sophistication. The co-assembly of the viral RNA and one (or a few) proteins to build a helical particle is one of the simplest arrangements, widely used by single stranded (ss) RNA and ssDNA viruses. While easy to build, these viral assemblies may be too rigid to adapt to different environments, and they have to be fully disassembled to allow viral genome replication and expression.

A more complex structural solution is offered by the enveloped RNA viruses, where a flexible nucleocapsid is formed by the interaction of the RNA and multiple copies of a protein (in a way resembling the simple solution mentioned above); but in this case, the nucleoprotein assembly is based on very different types of RNA-protein contacts, and may adopt different structural conformations to carry out the transcription and replication of the genetic message. Proper protection to transfer this complex from cell to cell is conferred by the inclusion of the nucleocapsid inside a lipidic envelope derived from the cellular membranes after its modification with viral proteins. The extreme in this type of strategy is offered by the more complex viruses (human immunodeficiency virus, adenoviruses, herpesviruses, vaccinia virus), where the nucleocapsid is enclosed by several layers or envelopes of different composition. Examples of these viruses are given in Chap. 11 and several other chapters of this book.

An alternative strategy, used by many viruses, is to incorporate the nucleic acid while the capsid is being formed in a co-assembly process. This solution is adopted by a large number of non-enveloped ssRNA viruses with icosahedral capsids. Icosahedral symmetry is a preferred geometrical solution to build a closed viral container (see Chap. 2). Although the assembly details vary from one virus type to another (see Chap. 10), it seems that the interaction of small segments of the RNA molecule and certain domains of the shell protein facilitates their mutual recognition, and their action as chaperones that assist each other to drive the co-assembly process.

The extreme rigidity and other structural properties of double-stranded (ds) nucleic acids impose severe restrictions to the possible ways to enclose these molecules within a protein container. Nevertheless, the evident evolutionary advantages of dsDNA as a genomic substrate have driven viruses to explore efficient solutions to encapsidate this type of DNA molecule, the best known example being many bacteriophages. These viruses assemble first a protein container (called prohead), in which dsDNA is later packaged using a complex machinery made of different viral proteins. This process requires an exquisite selection procedure to encapsidate the right (viral) DNA, followed by an energy-consuming process that translocates the DNA up to quasi-crystalline densities inside the virus shell. These viruses can be considered as a paradigm of efficient

DNA aggregation: there is no other biological example for such an ordered and condensed DNA conformation, an order that plays a critical role in the release of the DNA during infection by these phages. Nevertheless, this packaging solution is expensive in terms of information (several viral proteins are involved), as well as in energy consumption (one ATP molecule is hydrolyzed by each two base pairs packaged), but the reward is worth the expense: These viruses exhibit the largest infection efficiency by far, as each viral particle is potentially able to infect a host bacteria.

In the following sections we will briefly review examples of viruses using different viral strategies for nucleic acid packaging, making emphasis in common underlying mechanisms and solutions selected by viral evolution.

12.2 Structural Features of the Packaged Nucleic Acids

The extended use of X-ray fiber and crystal diffraction and electron microscopy has provided a great amount of information on the structure of viral particles (see Chaps. 3, 4, and 7). Due to their geometrical shapes and the existence of symmetries (see Chap. 2), these and other structural methods have provided a comprehensive insight into the way the protein components of viral capsids are organized at the molecular or even atomic level. Nevertheless, the same cannot be said for the nucleic acid component of virions, as it is usually far less ordered than the protein counterpart, and in most cases the way the nucleic acid is organized in each viral particle is not identical to those in others. This section reviews the main features of the way the nucleic acids interact with the protein capsids for protection, mobility and functionality.

12.2.1 RNA Viruses

Tobacco mosaic virus (TMV), a representative positive ssRNA virus with helical symmetry has provided a most successful model for biophysical studies for more than 100 years. The structure of this rigid rod-shaped virion was solved at near-atomic resolution using X-ray fiber diffraction and later by electron cryo-microscopy [2, 3]. The virus is organized as a stack of identical subunits of a single capsid protein that follows a helical path with 17 subunits per ring, leaving an internal cavity (Fig. 12.1a). The ssRNA molecule is located near this inner region, with three nucleotides binding to each of the capsid subunits.

The type of protection for the RNA provided by TMV-like viral particles is limited in functional terms due to their rigid nature. A more flexible solution is adopted by negative ssRNA viruses. The Mononegavirales [8] are characterized by the assembly of a helical-symmetry based nucleoprotein complex (NPC) essentially made by the ssRNA and a specific protein (the nucleocapsid protein). Unlike the

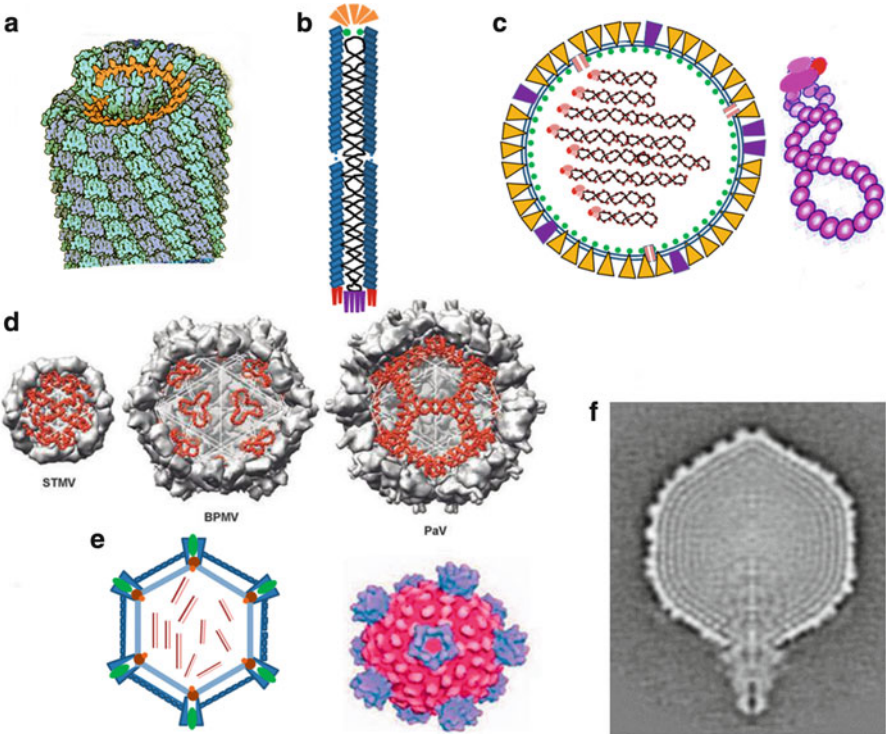


Fig. 12.1 Structural alternatives for nucleic acid encapsidation in viruses. (**a**) Structure of TMV. The viral RNA is colored *orange* (Adapted from the RCSB PDB. With permission). (**b**) Bacteriophage M13. The viral DNA is depicted as a superhelical molecule inside the virus shell. (**c**) Influenza virus. The different structural components are indicated (Adapted from [4]. With permission). The viral ribonucleoprotein complex is shown as a beaded string with the polymerase at one end. (**d**) Structures of STMV, BPMV and PaV particles sectioned to show the partial organization of the viral RNA molecule (*red*) (Reproduced from [5]. With permission). (**e**) Reovirus and the structure of the internal capsid solved by electron cryo-microscopy (Reproduced from [6]. With permission). (**f**) Section of the mature phage T7 structure obtained by electron cryo-microscopy. The DNA is shown as concentric punctuated layers inside the capsid (Reproduced from [7]. With permission)

helical TMV particle, NPCs have a remarkable intrinsic flexibility [9]. Instead of resembling a rigid rod, NPCs have a herring-bone aspect by electron microscopy, revealing different conformations with helical symmetry. These nucleoproteins are enclosed into a near-spherical or elongated lipidic envelope (integrating viral specific proteins) to yield the infectious viral particle (Fig. 12.1c). Examples of these viruses include the *Rhabdoviridae* (like the bullet-shaped vesicular stomatitis virus, or the rabies virus), the *Paramyxoviridae* (respiratory syncytial virus, Sendai virus or measles virus), and the *Orthomyxoviridae* (influenza virus). A general characteristic of all these viruses is that the NPCs (either one or, in the case of viruses with segmented genomes, several NPCs containing a ssRNA molecule each,

as in influenza virus) are released inside the cell and used as substrates for RNA replication and transcription. This means that the RNA-protein assembly must be flexible enough to allow the RNA polymerase (which also forms part of the NPC) to act on the viral RNA substrate without complete disassembly of the nucleoprotein.

Although the viral nucleocapsid proteins of different viruses (N-proteins) show a rather low sequence identity, electron microscopy and X-ray diffraction studies have revealed that NPCs share a basic structural organization that is relatively conserved. The NPC helix has a pitch of around 6–7 nm and 10–13 subunits per turn. Each protein subunit is associated with nine nucleotides, yielding a threading of the proteins into a super-helical complex. The RNA moiety is located in a conserved RNA-binding cavity formed between the carboxyl and amino terminal domains of the protein. This region presents a number of positively charged residues that can interact with the RNA phosphate groups, resulting in the exposure of certain bases of the RNA in the outer face of the NPC, fully available for functional recognition of the RNA polymerase. Three dimensional reconstructions of different NPCs based on electron microscopy studies showed differences in the overall superhelicity and different parameters of the basic helical assembly, and revealed that these complexes have an intrinsic flexibility that accommodates the conformations required in each of the different stages of the viral life cycle, either inside the cell or when packaged inside the viral envelope (reviewed in [10]).

A different solution for the successful storage of the viral genetic information based on ssRNA molecules is found in many icosahedral viruses [5]. Icosahedral symmetry-based capsids are found in about half the virus families. Among them, there is an important group containing positive ssRNA that includes human pathogens such as poliovirus, rhinovirus and hepatitis A, as well as many insect and plant viruses (*e.g.*, flock house virus (FHV) and satellite tobacco mosaic virus (STMV)). In other ssRNA viruses the icosahedral capsid is enclosed by an outer lipidic envelope, as in the *Togaviridae* and *Flaviviridae*. In many of these viruses, the RNA, besides carrying the genetic information, plays an important role in the assembly of the viral particle and in the definition of the shape and size of the capsid. Most of the knowledge we have on the function of the viral RNA in the assembly and life cycle of these viruses comes from biochemical and genetic studies. Knowledge of the RNA structure within viral particles has been largely prevented by the fact that all or a high percentage of the viral RNA is not sufficiently ordered inside the virion. Both X-ray crystallography and three dimensional reconstructions from electron cryo-microscopy images are based on the presence of order and symmetry in identical particles and, thus, only those parts of the RNA sufficiently ordered can be solved with a resolution comparable to that obtained for the capsid. In general, these viruses present ordered RNA density in contact with the capsid inner wall corresponding to around 13 % of the total RNA in the case of FHV, 20 % in bean pod mottle virus (BPMV), or 44 % in STMV. In all cases, the packaged RNA shows a dramatic increase of secondary structure compared to the cytoplasmic free RNA, including the formation of intrachain duplex RNA stems (up to 80 % of the total RNA is found in stem-loop structures within the

STMV capsid), as well as non-covalent base-pairing between RNA segments in those cases of multipartite RNA genomes.

The combination of X-ray diffraction results and electron microscopy reconstructions (see Chap. 7) has provided some insights on the organization of the RNA inside the icosahedral capsids of these viruses. As mentioned above, 30 dsRNA ordered segments are visible in STMV; they are 9 bp long each and follow the twofold axes of the icosahedral shell. By contrast, in BPMV the ordered RNA segments are located at the threefold axes (Fig. 12.1d), and in FHV (with a bipartite RNA genome) ordered double stranded RNA stretches, 10 bp long, are also observed along the capsid twofold axes. Interestingly enough, the location of these RNA moieties close to the interfaces between neighboring capsid subunits allow them to act as part of molecular switches to generate two types of quasi-equivalent intersubunit contacts required for the formation of either planar or more wedge-shaped interfaces along the twofold icosahedral axes (see Chap. 2). In certain cases it has been possible to see the overall topology of most of the RNA within the virion, as in the case of FHV and in the related Pariacoto virus (PaV), where the encapsidated RNA is organized as a dodecahedral cage of double-stranded dsRNA intrachain stems closely following the capsid twofold axes of the icosahedral shell (Fig. 12.1d) [5, 11].

There is evidence that, in some ssRNA viruses, the RNA plays a specific role in the promotion and direction of viral assembly [11, 12]. The fact that the RNA is folded in precise structures near the capsid and interacts with the capsid proteins supports a condensing role of the RNA in the nucleation and first assembly steps. There is also evidence that the capsid proteins could work as chaperones to facilitate the folding of the RNA in a subset of structures compatible with capsid geometry. Most probably, the mutual interactions between capsid proteins and viral RNA interplay to yield the virion definitive structure [5]. In addition, the existence of specific secondary structures in the RNA seems to play a fundamental role as packaging signals for the specific encapsidation of the viral RNA.

Another type of packaging solution is found in dsRNA viruses, which includes mammalian orthoreoviruses and certain fungal viruses and bacteriophages. All of them share many functional and structural characteristics, like the presence of concentric icosahedral capsids [13]. The study of the structure of icosahedral dsRNA viruses is an excellent example of the advantages of the use of hybrid methods: Combining X-ray diffraction of structural components and electron microscopy reconstructions of complete virions allows to obtain quasi-atomic maps of the different virus assemblies (see Chap. 7). The best studied representatives of this type of viruses are the *Reoviridae* family [14] and the *Cystoviridae* bacteriophage Φ6, which present a multipartite genome (three segments in Φ6 and 10–13 segments in reoviruses), together with several concentric icosahedral capsids (normally 2), and an outer lipidic envelope in the case of the *Cystoviridae*.

A common trend in all these viruses, except for the *Birnaviridae*, is the presence of an inner icosahedral capsid built by 120 protein subunits [15]; this capsid might be considered as a functional counterpart of the helical NPCs described above, but

in this case the capsid plays an active role in the viral RNA metabolism, namely in RNA replication, transcription and release from the capsid. The inner capsid in these viruses presents specific turret-like structures at the fivefold vertices (Fig. 12.1e). These assemblies, which contain the viral RNA polymerase, are involved in the generation and extrusion of the viral messenger RNA to the cell. In fact, this capsid maintains the dsRNA enclosed inside during the whole viral cycle, thus preventing a cellular response to the presence of the dsRNA molecules. Also, there is solid evidence that the RNA segments of the multipartite genome of these viruses are actually packaged into preformed inner core capsids, although in this case only a single vertex would be involved (see below).

All these fascinating functions carried out by the inner core have to be performed under quite extreme conditions: The dsRNA segments are densely packaged inside the capsid building concentric layers with spacings between 2.5 and 3.0 nm [16]. This high RNA density inside the capsid, together with the proposed RNA packaging mechanism, resembles the case of the dsDNA bacteriophages, which is also discussed below.

12.2.2 DNA Viruses

The *Inoviridae* are ssDNA viruses which present some of the simplest helical capsids. They infect bacteria and some of them, including M13, Fd and related phages and PF1, have been the subject of detailed studies. M13 has been widely used as a substrate for biotechnological approaches [17] (see also Chap. 22). The capsid protein has a cylindrical shape with the negatively charged amino terminal segment on the outside, and the positively charged carboxyl terminal segment inside, lining a channel where the DNA is enclosed (Fig. 12.1b). X-ray fiber diffraction studies of phage PF1 have shown that the DNA follows the helical parameters of the protein helix assembly (1.6 nm pitch), with lysine and arginine residues inserting between the DNA bases, to stabilize the phosphate charges [18]. Although the details of the protein-nucleic acid are specific in each case, the overall assembly strategy of these viruses closely resembles that of the ssRNA helical viruses mentioned above.

As in the case of the RNA viruses, the most widely preferred geometry of ssDNA viruses is the spherical shape using icosahedral symmetry. Most of these viruses infect mammals and other vertebrates (parvoviruses), while there are several examples infecting plants (geminiviruses) and bacteria (phage ΦX174). The resolution of the structure of canine parvovirus showed that 13 % of the total viral DNA is organized in defined structural stretches. These stretches are 11 nucleotides long and interact with capsid protein pockets at 60 icosahedral symmetry-related positions [19]. There is a low level of sequence specificity in these DNA-protein contacts, and there are no hints of DNA-binding motifs in the capsid subunits. In fact, DNA-binding domains are generated after capsid assembly, thus representing a clear mean to stabilize correct assemblies. A singular aspect in parvovirus is the

fact that the interaction of the DNA with the capsid protein is based on the existence of a series of well conserved polar residues in the inner face of the capsid, instead of the more common interaction of the nucleic acid with basic residues found in many other viruses.

The dsDNA viruses provide some of the best known examples of sophisticated strategies to package the nucleic acid in a most efficient way. Most of these viruses have a single linear dsDNA molecule enclosed within an icosahedral capsid, although there are examples of filamentous dsDNA viruses (*Lipothrixviridae*), and others show more complex architectures that include lipidic envelopes (herpesviruses, poxviruses). dsDNA viruses as a group can infect every cell type, from bacteria (some bacteriophage families), invertebrates (*Baculoviridae*, *Iridoviridae*) and vertebrates (*Papovaviridae*, *Adenoviridae*). The best known examples of this type of viruses include the caudovirales, a widely extended bacteriophage order characterized by their icosahedral capsid attached to a tail which is instrumental for the phage-host interaction [20].

Caudovirales present a common maturation strategy, building first a proteinaceous prohead, which is later filled up with the dsDNA (for a review see Ref. [7]). The physical properties of the dsDNA molecule clearly impose severe constrains on the way it can be packaged inside a spherical (icosahedral) container. The persistence length of the dsDNA (around 50 nm) is related to a minimum diameter in the capsid of these viruses: most of them have diameters from 50 to 400 nm, and they present a complex DNA packaging machinery (see Sect. 12.4) which sits in a unique icosahedral vertex of the viral capsid (the portal). This machinery has to deal with the formidable task to fill up the viral capsid with dsDNA up to quasi-crystalline densities (more than 500 mg/ml). Electron cryo-microscopy has revealed that the DNA is organized in layers following a traverse spool [21] (Fig. 12.1f). The distance between adjacent layers is 2.5–2.7 nm, suggesting that the DNA is tightly arranged in a close hexagonal packaging. Depending on the capsid geometry, the DNA seems to adopt a different topology: fully icosahedral capsids (T7, ε15, P22) shows the coaxial spool geometry, while in elongated icosahedral variants (prolate shells in T4, Φ29) the DNA strands adopt an orientation more parallel to the longitudinal axis of the viral capsid [22]. In all these topologies the structure of the outer layers is always better defined than the inner layers: it seems clear that the inner DNA region must present sharp geometrical discontinuities (bends) to allow its accommodation in the very restricted available space. In this context, the presence of inner capsid proteins must play an important role. These components can be individual small proteins (probably instrumental in shielding the electrostatic repulsion between phosphates of neighboring DNA segments); alternatively, they can form a complex inner core connected to the portal vertex (as in T7, ε15 and myovirus) (Fig. 12.1f). The presence of this structure along the DNA translocating path appears to be related with both the way the DNA is accommodated in the inner shell surface and the ejection process [7].

An important factor to be considered in these viruses is that this highly condensed DNA inside the viral head has to be also efficiently injected into the

host bacterium. This imposes important constraints to the possible topological arrangements of the DNA, especially taking into account that although the packaging and release of DNA are performed along the same capsid vertex (the portal), the respective molecular mechanisms in which they are based are different. While the packaging reaction consumes cellular ATP to develop up to 50 pN forces (see Sect. 12.5.1 and Chap. 9), the ejection of DNA initially depends on structural changes in the tail apparatus that triggers the release of the potential energy stored in the packaged viral head.

Different studies by electron cryo-microscopy have shown that the DNA packaging pathway in these viruses follows an initial stage where the DNA is pushed inside the capsid in a poorly ordered way. Only when the percentage of DNA is around 60–70 % of the total to be packaged in phages as T3 or Φ29, the layers of DNA start to organize, probably from the layer in contact with the inner face of the capsid shell [23, 24]. At higher DNA concentrations the multiple layer topology is generated until the acquisition of the final coiled spool structure.

12.3 Reorganization of the Viral Capsid During Nucleic Acid Packaging

The viruses that actively package their genetic material (with energy consumption) start their morphogenetic pathway by forming an empty shell (prohead) that will be subsequently filled up with the nucleic acid by a molecular motor. This packaging machinery translocates the viral genome to a very high concentration inside the head (see Sect. 12.2). Nucleic acid encapsulation correlates with a change in capsid morphology in a process known as virus maturation (see Chap. 13). These conformational changes have been best characterized for dsRNA and dsDNA bacteriophages. Although associated in many cases to nucleic acid packaging, maturation may occur before, during and/or after packaging, and is described in detail in Chap. 13.

12.3.1 Conformational Changes During RNA Packaging

Active packaging in RNA viruses has only been described for Φ6 *Pseudomonas* bacteriophage [25]. Φ6 is composed by three concentric layers that enclose a segmented dsRNA genome (Fig. 12.2a) [27]. As it was previously mentioned (Sect. 12.2.1) this bacteriophage shares functional and structural characteristics (common capsid fold and common head symmetries) with other members of the *Reoviridae* family (like bluetongue or Rotavirus). Nevertheless it remains unclear whether these viruses share as well a common packaging mechanism [28].

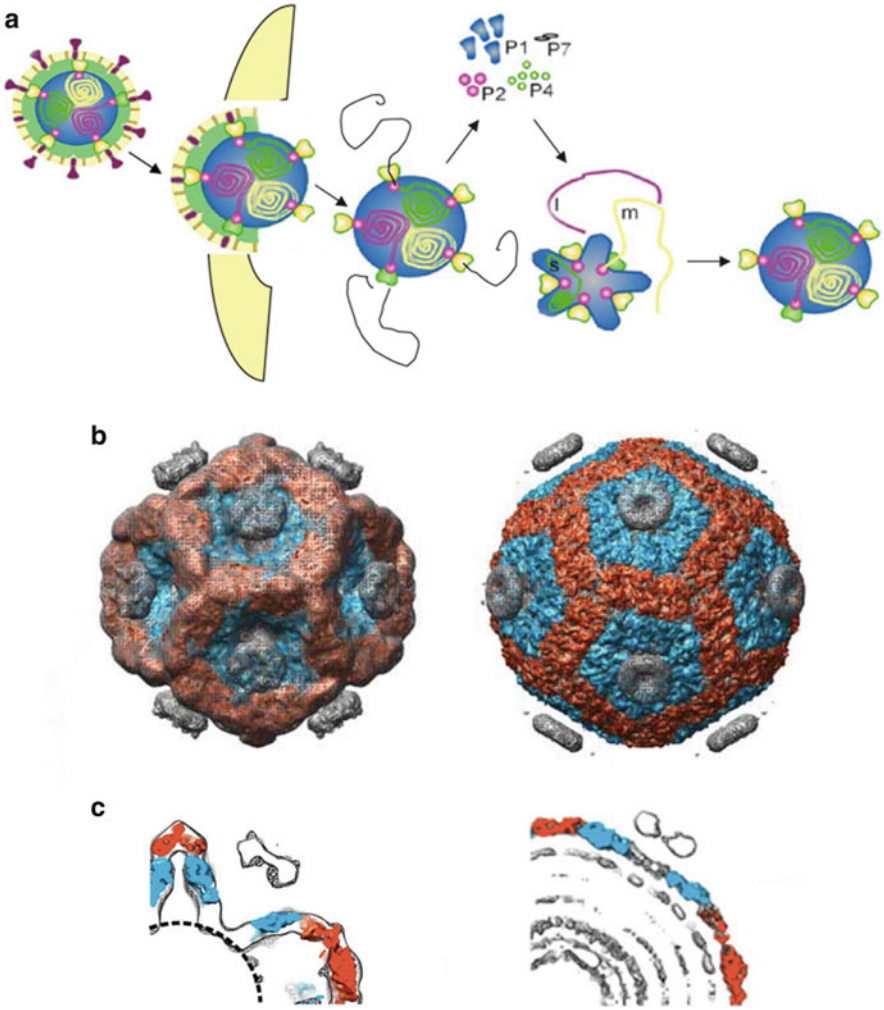


Fig. 12.2 Conformational changes during dsRNA packaging. (**a**) Morphogenetic pathway of Φ6 bacteriophage showing (from *left* to *right*): viral entry inside bacteria, extrusion of mRNA templates, synthesis of viral proteins, procapsid assembly and RNA packaging. (**b**) *Left*, model of Φ6 prohead before RNA packaging showing cup-like structures at the fivefold symmetry axes. *Right*, model of the capsid after RNA packaging. P1 (capsid protein) class A monomers are shown in *blue* and P1 class B monomers are shown in *red*. (**c**) Sections of capsid structural models before (*left*) and after (*right*) RNA packaging, showing the conformational change that takes place in the P1 monomers; color coding for the P1 monomers is the same used in (b) (Figures in panels (b) and (c) are reproduced from [26]. With permission)

In the case of the *Cystoviridae* family, the multi-layered capsid is also enveloped by a lipid bilayer. The two first layers are lost during viral entry and the innermost shell is delivered inside the cytoplasm activating RNA synthesis (Fig. 12.2b) [28]. During assembly cystoviruses form an empty procapsid, whose main component is protein P1 [25]. The procapsid shows an icosahedral shape composed by cup-like structures assembled on the fivefold vertex [28]. The fivefold vertex of the capsid is occupied by the RNA-dependent polymerase (P2), the packaging NTPase (P4) and the packaging factor (P7). The structure of the procapsid differs from the RNA-filled head. During RNA packaging the prohead structure expands sequentially to a spherical shaped structure with turrets projecting at the fivefold vertex. This conformational change has been suggested to play a role orchestrating RNA segment packaging order by sequentially exposing different genome interaction regions [28]. During this capsid rearrangement the internal volume of the capsid is increased by 2.4-fold [26].

The inner capsid is organized in a T2 lattice and the asymmetric unit is composed by a dimer of two P1 monomers, A and B. The class A monomers surround the fivefold axes, while the class B monomers assemble the pentamers together (Fig. 12.2c). Capsid expansion can be explained by local conformational changes at the P1 monomer interfaces. The most dramatic changes are produced at the A-P1 monomers interface that changes from an angular conformation to a nearly flat position (Fig. 12.2c) [26].

12.3.2 Expansion and Reinforcement of Caudoviral Procapsids During DNA Packaging

As we have mentioned above, complex dsDNA bacteriophages (caudovirales) are among the best known viruses that package their dsDNA into preformed viral proheads. These proheads (Fig. 12.3a) are composed by an outer shell, an inner scaffold, and in some cases a core-like structure, as discussed in Sect. 12.2 [20]. There is also a complex machinery located at the portal vertex of the prohead (see Sect. 12.4), which is directly involved in the translocation of the DNA inside the preformed protein container [30]. The onset of the DNA packaging is correlated with a number of changes in these proheads: there is a massive rearrangement of the shell subunits, leading to the expansion of the head diameter by about 10 %, the scaffold is disassembled and their subunits are either proteolyzed or recycled in new prohead assemblies, and the core components undergo also structural transformations [31]. In several viruses there is also incorporation of new proteins to the viral shell and/or chemical modification of shell proteins (cleavage, crosslinking). It is worth emphasizing that all these rearrangements have to be carried out in situ, maintaining the structural integrity of the viral particle while the

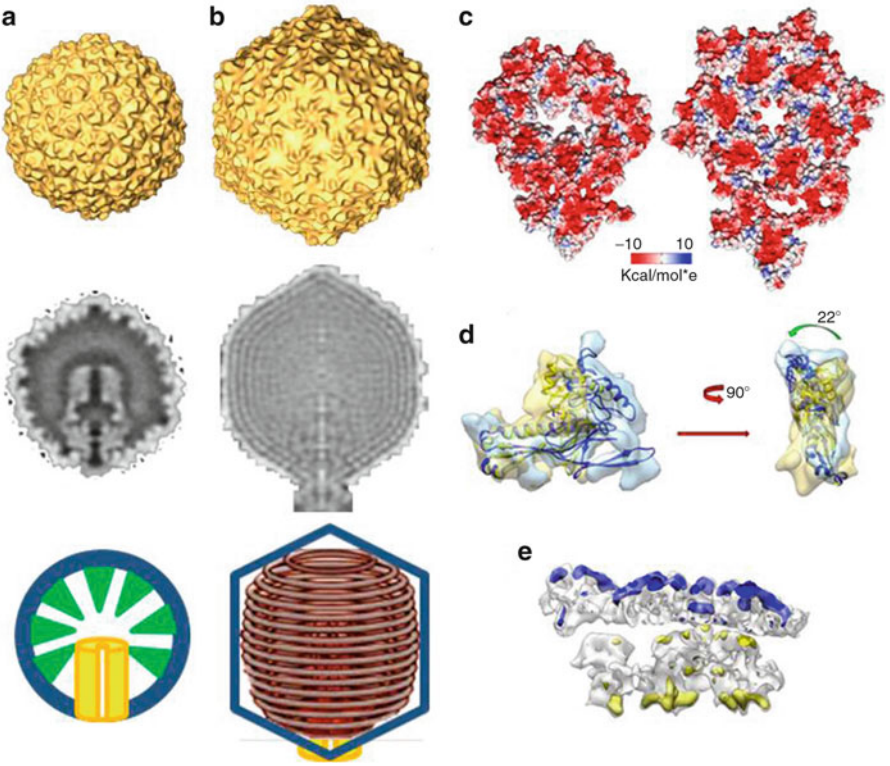


Fig. 12.3 Reorganization of *Caudovirales* capsids during DNA packaging and maturation. (**a**) Structure of the prohead of phage T7: electron cryo-microscopy reconstruction (*top*), section revealing the core and scaffolding components (*middle*) and schematic outline of the basic structural components (*bottom*). (**b**) Structure of the mature viral capsid of phage T7. The capsid is more angular than the prohead (*top*), the DNA and the core form a tight complex inside the capsid (*middle*), with the ordered DNA tightly coiled around the core (*cartoon at bottom*). (**c**) Electrostatic potential in the inner surfaces of the prohead (*left*) and mature head (*right*) of phage T7. (**d**) Two views of the relative rearrangements of the domains of the T7 main shell protein subunit when comparing the prohead (*yellow*) and mature head (*blue*). (**e**) Movement of the N-terminal domain of the main shell subunit of phage T7 from the inner side of the prohead shell (*yellow*) to the outer side of the mature head shell (*blue*) (Figures (c), (d) and (e) reproduced from [29]. With permission)

DNA molecule is actively being incorporated inside the shell interior. This reflects the exquisite architectural design of these proteins containers. The consequences of the structural rearrangements are fundamental for the packaging reaction: The viral capsid volume becomes almost 50 % bigger, the scaffolding protein is released and leaves the capsid, and a new inter-subunit interaction pattern is created (Fig. 12.3b). This reorganization makes more stable the viral shell and, at the same time, creates

a new inner shell surface (Fig. 12.3c): while the inner face of the prohead shell is quite uniformly electropositive, the mature head inner surface presents negatively charged domains that might be important for the interaction of the first layers of DNA to start the organization of the packaged DNA inside the capsid [29].

All these changes are closely related to the incorporation of the DNA, as the release of the scaffolding and the increment of the capsid volume facilitate the entrance of larger DNA molecules up to high concentrations. At the same time, the shell subunit rearrangement, based on local tilting, rotation and radial translations of the subunit domains (Fig. 12.3d), generates an increment in the contact areas between subunits, thus increasing the stability of the particle [29]. This increment might be further reinforced in different viral systems by either the incorporation of new, accessory proteins (as in T4, λ and ε15), or by covalent inter-subunit cross-linking (HK97) (reviewed in [32]). An additional aspect to be considered is the fact that during the shell reorganization, the shell protein domain that is involved in the putative contacts with the scaffolding protein changes its position from the inner side of the shell to the outside (Fig. 12.3e). This rearrangement not only facilitates (and probably induces) the release of the scaffolding protein, but also prevents the eventual competition between the DNA molecule and the scaffolding protein subunits for the interaction with the inner side of the shell. The final result is a more stable capsid, securely enclosing the DNA, and ready to incorporate the tail complex to produce the final infective viral particle.

12.4 Components of the Packaging Machinery

The packaging machinery is a molecular motor that transforms nucleoside triphosphate (NTP) hydrolysis energy into mechanical work, leading to nucleic acid translocation inside the viral head. These motors have evolved from the simplest ones present in dsRNA viruses to very complicated transient multi-complexes present in dsDNA bacteriophages.

12.4.1 dsRNA Packaging

Viruses containing a dsRNA genome present a unique packaging machinery. The existence of RNA as genomic material forces them to use a specialized viral RNA polymerase to replicate and generate its mRNA transcripts [27, 33]. To avoid degradation by host RNAses the viral polymerase operates inside the intact viral apparatus. The strategy of these viruses consists in replicating their RNA as it is translocated inside the viral head, and it requires the coordination of two viral complexes, the polymerase and the RNA translocase, both sitting at the fivefold vertex (Figs. 12.2 and 12.4a and Sect. 12.3.1). RNA packaging has been only characterized for dsRNA phages belonging to the *Cystoviridae* family [25, 27],

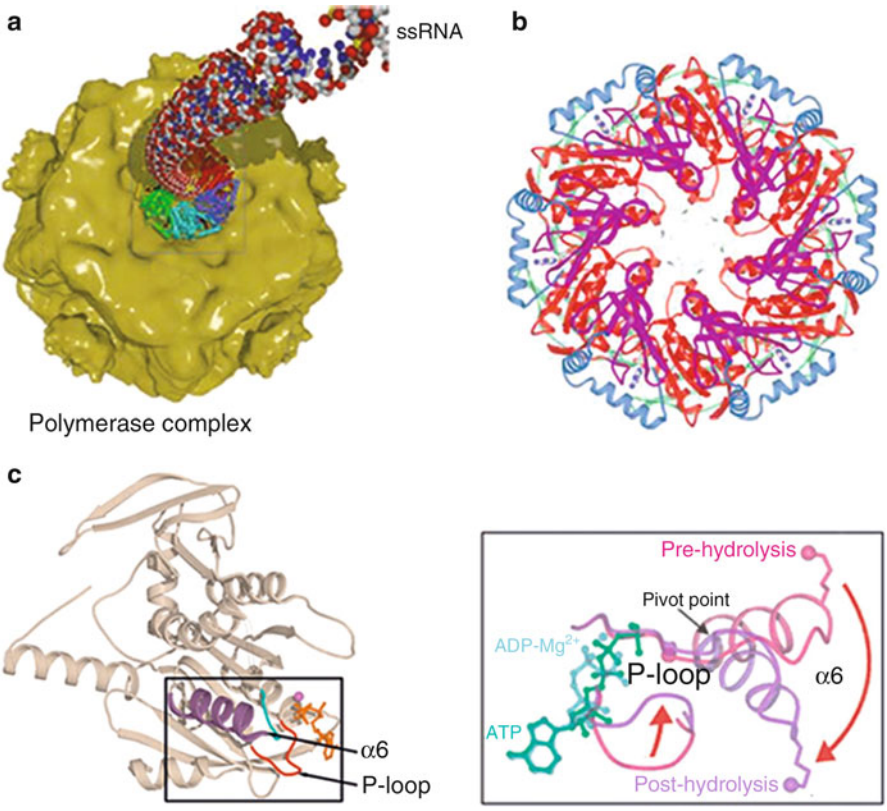


Fig. 12.4 P4 the packaging motor of Φ12 bacteriophage. (**a**) Model of the polymerase complex of Φ12 bacteriophage showing the P4 position at the fivefold vertex of the capsid translocating a ssRNA molecule (Reproduced from [27]. With permission). (**b**) Model of the atomic structure of the P4 hexamer (Reproduced from [34]. With permission). (**c**) *Left*, cartoon representation of the atomic model of one P4 monomer before ATP hydrolysis. *Right*, close-up of the region highlighted in the *left panel* showing the conformational change that takes place after ATP hydrolysis (Adapted from [27]. With permission)

and it is not clear at the moment whether this mechanism is also shared with other segmented dsRNA viruses belonging to the *Reoviridae* family [35].

Viral RNA translocases are structural proteins that are also present in the mature virus. They build the channel for ssRNA entry during packaging and exit of the messenger RNA (mRNA) transcript during the first steps of viral infection (Fig. 12.4a). RNA translocating systems are simple packaging machines as they are only constituted by one protein. P4, the packaging motor of bacteriophage Φ12, is an NTPase (it does not have ATP specificity), and it is the only one whose structure is known up to date [27]. Nevertheless this bacteriophage presents homology with other virus belonging to the *Cystoviridae* family (Φ6–Φ13). P4 shares important similarities with other helicases: the sixfold symmetry and the

characteristic Rossmann fold, consisting in a twisted eight-stranded β-sheet surrounded by five helices responsible for nucleotide binding (Fig. 12.4b) [27]. The crystal structure of P4 shows that this protein presents a conserved central core that, together with the C-terminal domain, constitutes the Rossmann fold. Despite lacking sequence homology, this nucleotide binding domain is structurally similar to the one of RecA, the T7 helicase and the bacterial conjugation protein TrW. The central channel is only wide enough to accommodate ssRNA. This channel is flanked by helix α6 and the L1 and L2 loops (Fig. 12.4b). Hydrogen-deuterium exchange and mutation of the lysines present in these loops prevents RNA packaging, indicating that these amino acids are implicated in RNA interaction and translocation [27]. Structural determination of two different nucleotide interacting protein intermediates suggested a mechanism for RNA translocation: The phosphate binding region (P-loop) changes from a down to up conformation after NTP hydrolysis, indicating that this modification could act as a molecular lever pumping the RNA inside the procapsid (Fig. 12.4c) [36].

The packaging mechanism can be subdivided into three steps: RNA recognition, RNA loading into the hexamer, and RNA translocation. The current model suggests that RNA recognition is mediated by conformational changes into the P1 capsid protein (see Sect. 12.3.1). RNA loading has been postulated to happen by ring opening, as is the case of the Rho protein. The interaction of the RNA molecule triggers NTPase activity leading to the translocation of the RNA molecule [36]. Once the molecule is inside the capsid, the viral polymerase generates the dsRNA molecules present in the mature viral particle.

12.4.2 dsDNA Packaging

As mentioned above, the DNA packaging mechanism is well conserved in dsDNA bacteriophages and in some animal viruses such as herpesviruses, showing common essential components. In this section we present a brief review of the most important of them to gather additional functional knowledge about this process.

The Portal Complex

The portal complex is located at a unique vertex of the prohead, and it comprises the connector (that builds a channel to accommodate the DNA) and the terminase, that is involved not only in the selection and processing of the DNA to be packaged but also in the ATP-driven DNA translocation. Although the icosahedral head has 12 pentameric vertices, the position of the connector and the transient interaction of the terminase is confined to a specific one. Moreover, in some viruses, such as T7 and ε15, there is a proteinaceous internal structure, called the core, also interacting with

Table 12.1 Viral components involved in dsDNA packaging

	Connector		Large terminase		Small terminase		dsDNA
Virus	Gene product	Mass (Kda)	Gene product	Mass (Kda)	Gene Product	Mass (Kda)	Length (Kb)
λ	gpB	59	gpA	73	gpNu	20	48.5
SPP1	gp6	57	gp2	49	gp1	21	45.9
T4	gp20	61	gp17	70	gp16	18	166
T7	gp8	59	gp19	67	gp18	20	40
P22	gp1	83	gp2	58	gp3	19	43.4
Φ29	gp10	35	gp16/pRNA	39/58	gp3	31	19.3
HSV1	UL6	74	UL15	81	UL28	85	152

this vertex. This core is not required for the prohead assembly, but it seems to be essential for infectivity and, as it has been revised in Sect. 12.2.2, it may facilitate the topological ordering of the dsDNA genome during packaging and/or release. Consequently, through a special vertex the DNA is both pushed by the portal complex and, once the virus is formed, ejected during the infection process [30, 37].

The function of the portal complex is defined not only by the communication among its own components but also by the interaction with the DNA substrate. The linear dsDNA molecule varies in length for each viral system: from 19.3 Kb in the case of Φ29 to 166 Kb in T4 (Table 12.1). Most of them (except Φ29 and some related phages) produce head-to-tail multimers, or concatemers, of DNA as a substrate for packaging. Concatemeric DNA is formed by genomic units linked by tandem repeats generated during the replication process. The recognition of the terminal sequences by the viral proteins ensures the specific packaging of the viral genome from that of the host. Moreover, the terminal repeats indicate the cutting region between genomic units, avoiding the loss of base pairs in the 5′ end, and ensuring the packaging of a single molecule. The specific recognition sequences, named cos or pac sequences, differ depending on the viral system. They can be identical in both extremes as in T7, T3 and λ; non-unique and repeated at the ends as in P22 and T4; or miscellaneous, with a protein covalently attached, as in Φ29. Besides this sequence recognition, two cleavages are required for the unit length packaging: the initiation cut (that generates the free end from which packaging starts) and the terminal cut (that delimitates the genome unit). Those cleavages can be either sequence specific, as in λ and T7, or sequence independent as in T4, P22 and SPP1. In the latter case, the connector acts as a termination sensor that produces the "headful-signal" when the genome is already packaged [38, 39].

The packaging proteins and their size, together with the genome length of several of the most abundant of those viruses (tailed bacteriophages) are presented in Table 12.1. Connector proteins and large terminases present wide variations in size, although the size of small terminases is more similar in different phages (except for HSV1, significantly higher). Even though the dsDNA molecule is always linear, its length differs considerably from one virus to another [38, 40].

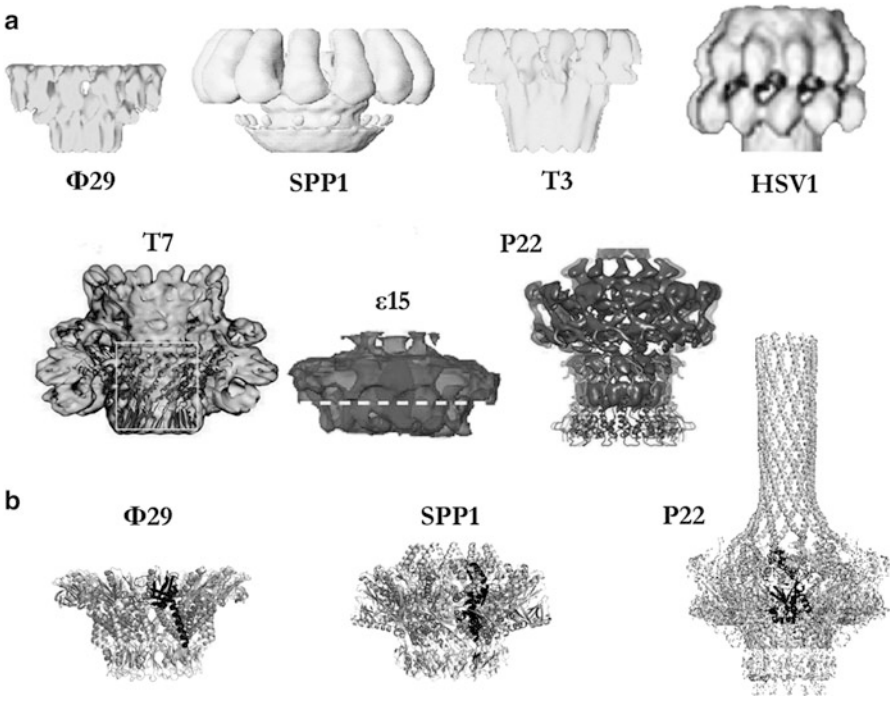


Fig. 12.5 Oligomeric connectors structures in dsDNA viruses. (**a**) Electron microscopy three-dimensional reconstructions of oligomeric connectors in different viruses. The structures share a toroidal morphology with wings protruding, a central channel (visible in the phage T7 connector) and a stem region in the basal part. The models present 12-fold rotational symmetry, except for the phage SPP1 connector model that was obtained for in its 13-mer form. (**b**) Atomic structures of the oligomeric connectors solved by X-ray crystallography. The conserved central domain is highlighted. PDB codes: phage Φ29 (1H5W), phage SPP1 (2JES) and phage P22 (3LJ5) (Reproduced from [41]. With permission)

The Connector

Although connector proteins do not share sequence homology, and they present large variations in size, they show a common architecture. Connectors are dodecameric assemblies with a toroidal morphology showing a conspicuous channel in the center. This ring shape is common to protein complexes involved in DNA translocation such as helicases, sliding clamps and certain exonucleases and topoisomerases, suggesting the implication of the channel in the DNA interaction. As shown in Fig. 12.5, several oligomeric connectors have been solved using electron microscopy and X-ray crystallography. These propeller-like structures can be divided in several domains (see Fig. 12.5a, b): the crown (apical part in HSV1, T7, ε15 and P22), the wing (central part, with maximum diameter) and the stem (the tight region in the base of the structures). The central domain is the most conserved, and it comprises two helices and an additional extended α-β domain

present in all bacteriophage connectors (see the highlighted ribbons in Fig. 12.5b). The other two domains, the crown and the stem, present variations in length and complexity (note the huge α-helical tube protruding from the crown of the P22 connector, in Fig. 12.5b). Considering the structural similarities among all connector proteins it is reasonable to assume that the smaller one would represent the minimum core to perform the main common functions while the supplemental domains, present in larger connectors, would correlate with the acquisition of additional features [41, 42].

All the connectors inserted in the viral head are built by 12 subunits, although the overproduction of this protein may induce the formation of different oligomeric forms. The symmetry mismatch between the dodecameric connector and the pentameric vertex plays an essential role during DNA translocation and it is also involved in the tail attachment to the portal. Moreover, the tenfold symmetry of the dsDNA might act as a mediator among the rotational symmetries of the portal complex and the vertex. The toroid oligomeric symmetry builds a channel with a diameter wide enough to fit the dsDNA molecule. The overall surface charge of the channel is mainly electronegative (especially at the entrance and exit) with rings of positive charges scattered all over the walls that may interact with the electronegative DNA molecule. Furthermore, SPP1 and Φ29 connectors present loops protruding inside the channel that may be involved in retaining the DNA inside the head [41].

The interactions between the connector and other viral components are fundamental for the portal functioning, as it has been shown by mutational analysis. The connector has been proposed to nucleate the prohead assembly by the interaction with the scaffolding proteins. The connector also interacts with the viral ATPase, serving as a docking point and modulating the packaging motor activity. In Φ29, an unusual case, the connector directly interacts with a specific virus-encoded small RNA molecule (pRNA), which encircles its narrow end acting as a bridge between the connector and the ATPase (see below in this Section). The P22 connector presents a helical barrel that would mediate the orderly DNA filling of the head and it also regulates the delivery pressure during DNA ejection. Additionally, the connector interacts with either the tail proteins or other proteins involved in the closure of the channel after DNA packaging. In fact, the connector has been proposed to be the sensor of the headful mechanism, defining the quantity of DNA to be packaged. Finally, to develop all its functions, during and after DNA packaging the connector probably undergoes conformational changes, which would not be necessarily irreversible as it was previously suggested [41, 43].

The Terminase Complex

The second component of the portal is the terminase complex, proposed to be the macromolecular motor that converts chemical energy from ATP hydrolysis into mechanical movement of DNA during phage morphogenesis. Beside the ATPase activity, most terminases also contain the endonuclease that cuts concatemeric

DNA into genome lengths. This nomenclature derives from the phage λ proteins that were first shown to be required for the formation of the termini of the packaged DNA. Terminases are generally hetero-oligomers built by a small protein involved in DNA recognition, and a large protein containing the ATPase and nuclease activities and a motif for docking at the portal vertex. Phage Φ29 is an unusual case, as its large terminase protein (gp16) does not have nuclease activity (the Φ29 DNA replicates as a unit length genome), and it functions in coordination with a small packaging RNA (pRNA) required to dock the gp16 onto the portal. Although this pRNA may be considered as part of the large terminase, it would be described separately below in this section. Even though there is no significant overall sequence similarity, terminase proteins from different phages contain well-conserved patches of amino acid sequences, or structural motifs that are required for packaging [38]. The structural information available about terminases, mainly obtained by X-ray crystallography, facilitates correlating their topology and interactions to their function along DNA packaging.

The large terminase subunit. Only a few atomic structures of large terminases have been obtained due to their flexible, conformationally heterogeneous nature (Fig. 12.6a). The first crystallographic structure presented was the gp17 amino terminal domain (Ndom) from RB49 phage, which shares 72 % sequence identity with its counterpart in phage T4. Afterwards, the carboxyl terminal domain (Cdom) of the same protein and the full-length crystallographic structure of gp17 from phage T4 (Fig. 12.6a, right panel) were solved. Recently, the atomic structures of the Cdom of both G2P and UL89, from phage SPP1 and human cytomegalovirus (HCMV) respectively, were obtained (Fig. 12.6a, left and central panel).

Sequence alignments show that the functional signatures of the ATPase domain of viral terminases are conserved and they are similar to those of the translocating monomeric SF2 helicases, restriction endonucleases, and protein translocases. Large terminases consist of two domains, an N-terminal ATPase domain that powers DNA translocation and a C-terminal nuclease domain that generates the termini of the viral genome. The Ndom of T4 consists of the classic nucleotide binding Rossmann fold (see Sect. 12.4.1). It contains the Walker A and B and the catalytic carboxylate, often found in proteins that bind and cleave ATP. Biochemical analysis of large terminases homologues from phages Φ29, λ, SPP1, P22, T3 and T4 confirm the ATPase activity of this domain. The nuclease Cdom presents an also conserved RNase H-like fold formed by seven β-sheets sandwiched between two clusters of α-helices. This basic fold (pointed by an asterisk in Fig. 12.6a) displays variations in length and shows almost no amino acid sequence identity [38].

The over expressed and purified large terminases from many phages exhibit different oligomeric states. T4 gp17, SPP1 G2P, P22 gp3 and HCMV UL89 exist essentially as monomers, while λ gpA oligomerizes in solution. Nevertheless, during the packaging machinery assembly, the terminase stoichiometry may be remodelled. The analysis of Φ29 and T4 large terminases bound to the portal vertex has shown electron densities that are consistent with five subunits being present. In the case of pRNA, six-prohead bound fluorescent molecules were reported, whereas

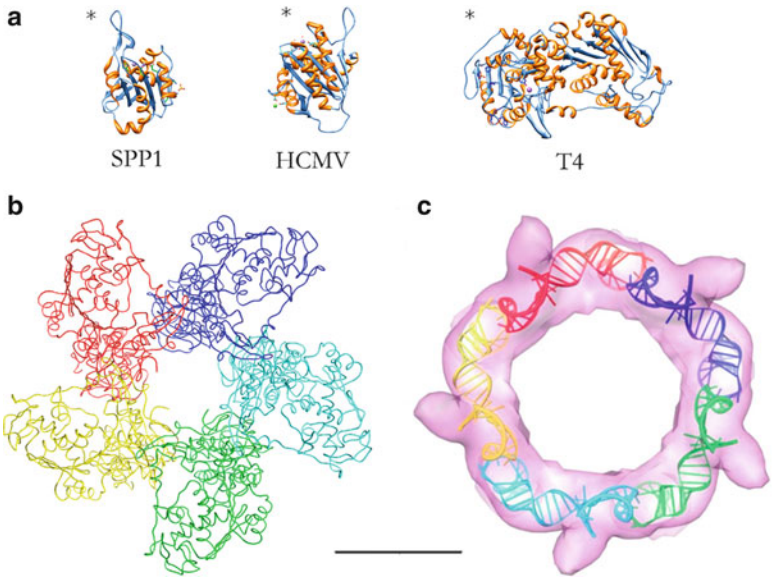


Fig. 12.6 Atomic structures of the large terminases and the pRNA solved by X-ray crystallography. (**a**) Large terminase monomers: nuclease domains of G2P and UL38 from phage SPP1 and HCMV, respectively; and complete structure of gp17, from phage T4. The nuclease domain (*asterisk*) shows an RNase H/integrase-like fold constituted by a bunch of parallel and anti-parallel β-sheets surrounded by several α-helix. (**b**) Frontal view of the atomic model of the pentameric gp17 terminase from phage T4. Each monomer is depicted in a different colour. (**c**) Flexible docking of the pRNA crystal structure into its electron cryo-microscopy pentameric envelope (Reproduced from [44]. With permission). The scale bar corresponds to 50 Å and refers to images in panels (b) and (c). Images in (a) and (b) were generated from atomic coordinates deposited in the PDB. Access numbers: phage SPP1 (2WC9), HCMV (3N4Q), phage T4 (monomer 3CPE, pentamer 3EZK) and phage Φ29 pRNA (3R4F)

recent studies support its interaction with the prohead forming a pentameric ring. Figure 12.6b represents the pentameric atomic model of T4 gp17, and the flexible fitting of the pRNA crystal structure into its electron cryo-microscopy pentameric envelope is presented in Fig. 12.6c. The oligomeric ring-like form, together with the diameter of the channel (40 Å in T4 gp17 and 82 Å in pRNA), allow the passage of the dsDNA molecule [45]. The terminase pentameric stoichiometry fits well with the tenfold DNA symmetry, but it mismatches with the 12-fold connector symmetry. The latter can be largely reconciled if there were interactions between the terminase and the pentameric vertex of the prohead.

During packaging the large terminase interacts with the connector (inserted in the prohead), with the DNA (first to translocate it and finally to cleave it) and with the DNA attached-small terminase. The binding to the connector-prohead is localized in the C-terminal region of most of the terminases, and it is based on charge-charge interactions. As suggested above, DNA is translocated through the channel of the large terminase, and the cleavage is carried out at a catalytic groove

of its Cdom. Finally, genetic and biochemical studies show that the interaction with the small subunit is localized on the Ndom of the large subunit [45, 46].

As the site for ATP binding and DNA cleavage resides in the large terminase, it has been proposed to undergo a conformational change in response to ATP hydrolysis to physically move the DNA. However, it has also been hypothesized that the energy from ATP cleavage is transmitted from the terminase to the connector, which in turn moves the DNA. The packaging models, together with the physical properties of the packaging process, will be revised in Sect. 12.5.

The small terminase subunit. Besides the DNA recognition function, small terminases perform a regulation of the large terminase enzymatic activities during DNA translocation. The small subunit is more variable in amino acid sequence that the other two motor proteins. Even though the Φ29 phage genome does not encode any small terminase, the gp3 protein primes its DNA developing an analogous function. In the last decade several structures of these proteins have been obtained (Fig. 12.7) using structural techniques such as nuclear magnetic resonance spectroscopy, electron cryo-microscopy and X-ray crystallography. Biochemical and structural data revealed how the domain organization is highly conserved among small terminases. The Ndom is the DNA binding domain, the central domain acts as the oligomerization domain and finally the C-terminal is in charge of the large terminase interaction. The N-terminal fragment of λ gpNu1 and the Shigella phage Sf6 g1p (Fig. 12.7a and c, respectively) share a similar winged helix-turn-helix fold. As shown in Figs. 12.7c, d, and e, the central domain is formed by two conserved α-helices and the Cdom share a characteristic β-barrel. The oligomerization of the small terminase produces multimer rings that vary in size from octamers to decamers (or larger). These data suggest that although oligomerization is important, the stoichiometry or a defined inner diameter does not appear to be strictly essential for its function. Alternatively, interactions with other components of the packaging machinery may select for a precise stoichiometry *in vivo*. Hence, the rings formed by the small terminases present variable channel diameters, from 10 to 24 Å in their narrow region. As a consequence, the wrap around model of DNA interaction has been increasingly favoured against the channel mediated one. Finally, despite the conserved overall structure of the small terminase, no conserved residues were located on the exterior surface indicating that the interaction with the large terminase or connector protein is mediated by the overall architecture of the portal complex and by the distinctive shape of the small terminase [45, 48].

Other Components: The pRNA

Bacteriophage Φ29 is an especial case because of the presence of a unique 174 nucleotides RNA molecule that forms part of its packaging motor together with the connector and the large terminase. Since this RNA was found to bind to proheads, it was named pRNA, and it is not needed to assemble proheads suggesting that pRNA attaches to proheads after capsid assembly. Its role is transitory and likely limited to

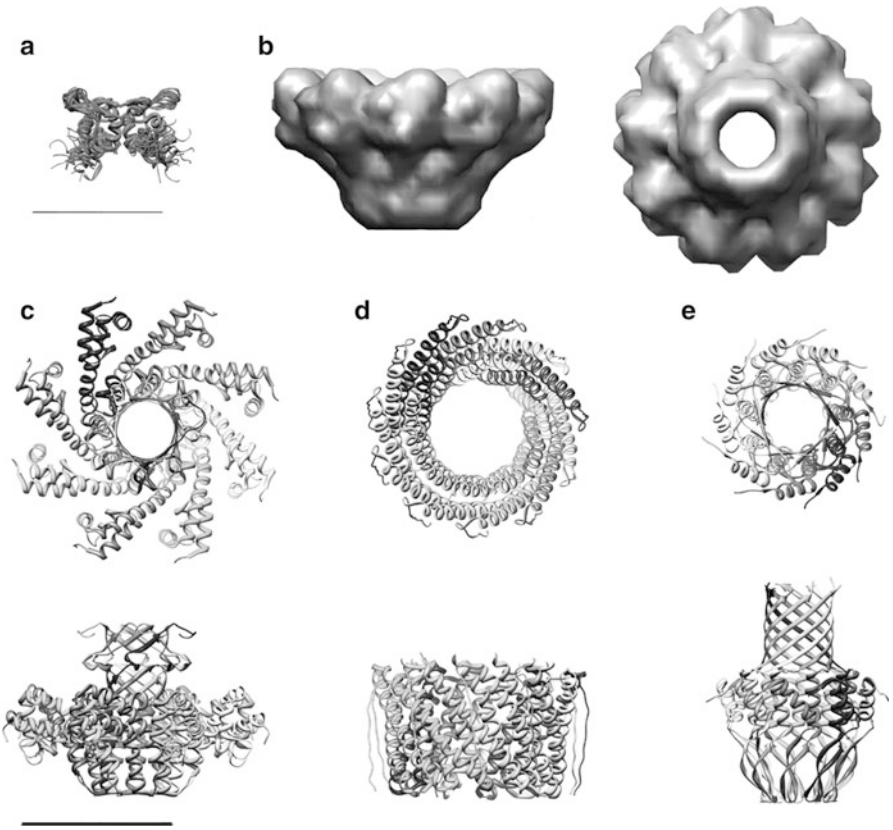


Fig. 12.7 Structures of the small terminases solved by different techniques. (**a**) DNA binding domain of the phage λ gpNu1 dimer solved by RMN. The scale bar corresponds to 25 Å. (**b**) Electron microscopy three dimensional reconstruction of the phage P22 gp3 nonamer at 18 Å-resolution (Reproduced from [47]. With permission). (**c**) Atomic structure of the octamer of g1p from the *Shigella flexneri* phage Sf6. (**d**) Crystal structure of the central region of the undecamer of gp16 from the 44RR phage (T4-like family). (**e**) Atomic structure of the nonamer of G1P from the *Bacillus subtilis* SF6 phage (SPP1-like family). In panels (c) to (e) frontal views (*upper panels*) and lateral views (*lower panels*) are shown. The scale bar corresponds to 50 Å and refers to sections (b), (c), (d) and (e). Images were generated from atomic coordinates deposited in the PDB. Access numbers: phage λ (1J9I), phage Sf6 (3HEF), phage 44RR (3TXQ) and phage SF6 (3ZQP)

DNA packaging, as pRNA is not present in mature Φ29. Thus, the connector, the pRNA and the ATPase form three concentric rings, and the DNA-gp3 is believed to be threaded through a continuous channel along their shared central axis and into the phage capsid [49].

The pRNA secondary structure consists of two domains: domain I comprises the first 117 bases and is separated from domain II by a 13 base single-stranded region. Although the full-length wild type is a 174-base transcript, a 120-base construct

encompassing domain I and lacking 54 nucleotides 3'-domain II is fully competent for packaging assembly *in vitro*. Despite the domain II function is not currently understood, its conservation in all Φ29 relatives suggests it may have an essential function *in vivo*. The domain I shows two functional modules: one necessary for prohead binding (helices C, D and E, and the three loops) and the other required for DNA packaging activity (the majority of the A helix), which provides the attachment site for the large terminase [50].

Formation of intermolecular pseudo-knots (also conserved in phage Φ29 relatives) between adjacent pRNAs would promote circular interactions resulting in an oligomeric ring-like structure. Nevertheless, the pRNA stoichiometry when bound to the prohead has been controversial. The hexameric state was supported by the analytical ultracentrifugation detection of both dimers and hexamers of pRNA [51]; and also by single particle fluorescence quenching experiments [52]. More recently, electron cryo-microscopy analysis definitely showed that bases 22–84 form a pentameric ring-like structure that binds to the prohead (see above and Fig. 12.6c), that five pRNA A-helices extend as spokes from this central ring, and that the terminase attaches to the distal end of those A-helix spokes [44]. The fitting on the electron cryo-microscopy envelope also revealed that although the pRNA was close to the connector, it made much more extensive contacts with the fivefold prohead vertex than with the dodecameric connector. This is in accordance with the effective binding of pRNA to connector-less proheads [50].

12.4.3 Other ATPase-Like Motors

Inside the cells there are proteins involved in the transport of nucleic acid between different compartments that have some resemblance to the viral DNA translocation motors. These proteins are ATPases sharing a RecA-like fold and assembled into propeller-like structures with a central channel that serves to transport the nucleic acid [53]. Some of these proteins play essential roles in chromosome segregation between two twin cells (as FtsK), or in bacterial conjugation to exchange genetic information between two different bacteria (as TrwB). FtsK translocates dsDNA and presents also a ring hexameric assembly. The structure is divided into two domains assembled through a central linker as in T4 ATPase (see Sect. 12.4). This structure suggests that during translocation the two domains would open like a jaw moving the DNA in an inchworm movement with a step size of 2 bp/ATP, similar to the one found in the viral DNA packaging motors in bulk experiments [54]. TrwB translocates ssDNA, and it is also an ATPase with a toroid structure assembled into a hexamer. Its atomic structure showed that it presents a RecA like fold. It has been suggested that ATP hydrolysis will trigger a conformational change allowing opening the central channel, which is too narrow in the solved atomic structure to translocate the dsDNA molecule [55].

12.5 Models for Nucleic Acid Packaging

The structural characterization of different components involved in viral nucleic acid packaging allowed proposing different models to understand the genome translocation process. Recent studies on single molecule experiments permit defining the mechanical characteristics of these motors (force generation, velocity, processivity and motor steps) using an optical tweezers setting (see Chap. 9). These experiments have been mostly carried for dsDNA bacteriophages. In this section we will first describe the properties of different machineries and then we will describe the mechanical models that have been proposed during the years. These models try to understand the coupling between ATP hydrolysis energy production and the conformational changes that lead to nucleic acid pumping inside the viral head.

12.5.1 Biophysical Properties of the Packaging Motor

The accurate measurement of viral genome packaging kinetics is difficult without the synchronization of different populations. This process is even harder for dsDNA motors where the motor only assembles transitory but it is not a structural part of the mature virus (see Sect. 12.4.2). While biochemical bulk assays can only determine the packaging efficiency by means of the total time to package the nucleic acid, single-molecule experiments allow collecting statistics of single events, and they have allowed observing for the first time pauses where the motor eventually disassembles from the packaged substrate (see Chap. 9) [56].

The optical tweezers setting usually consists in a viral prohead assembled with the packaging ATPase in the presence of an ATP analogue, then a microsphere carrying DNA is introduced and the variations in the nucleic acid extension values are carefully followed while different forces are applied to the system [45] (see Chap. 9 for technical details). These experiments have revealed the speed of the packaging motor for three different dsDNA bacteriophages: T4, Φ29 and λ [45, 56]. The speed values turned out to vary from 180 bp/s for Φ29 to 1,800 bp/s for T4, on average. The speed of the motor seem to be closely related to the length of the phage genome (see Table 12.1 and Sect. 12.4.2), thus phages with larger genomes have adapted to faster velocities in order to complete DNA encapsulation within 2–3 min from the total viral cycle of 20–30 min. It has also been observed that the speed decreased at late stages of packaging, when the internal pressure of the capsid increases. For example, λ packaging speed drops threefold when 90 % of the genome is inside the capsid.

Another biophysical value determined is the step size, or the translocation length of the motor after one ATP molecule hydrolysis. In bulk assays this value was determined to be 2 bp/ATP. Single molecule experiments have shown however, that he step size cannot be taken as a fixed value, as it can vary according to DNA symmetry and filling of the capsid. High resolution optical tweezers recently

showed that the motor translocates in four 2.5 bp steps (see Chap. 9). This strongly suggests a high coordination of the subunits, as the terminase is composed of five subunits and only four of them bind ATP before the hydrolysis and translocation are triggered [56].

Optical tweezers have also revealed that packaging motors are able to generate forces up to 50-60 pN at low capsid filling (see Chap. 9). These forces are 20–25 times that of myosine II, implicated in muscle contraction. Such high values may be essential to overcome internal capsid pressure, electrostatic repulsive forces and DNA bending (see [45] and Chap. 19).

The DNA packaging machinery is considered as a motor able to transform the chemical energy of ATP hydrolysis in a mechanical work consisting in nucleic acid translocation. One of the most interesting features to understand the mechanism of nucleic acid packaging is the coupling of these two events. During nucleic acid translocation the motor must interact with DNA and translocate the step size, release the DNA and set the process to zero to restart a new cycle. The ATP reaction can be divided in several steps consisting in ATP docking, nucleophilic attack of the γ-phosphate, and release of reaction subproducts. Recent studies with optical tweezers suggest that Φ29 engages DNA after ATP binding and that DNA is translocated right before ADP release [56].

12.5.2 dsRNA Packaging Models

Considering the large differences described between DNA and RNA packaging motors (see Sect. 12.4) it would be reasonable to think that RNA packaging machineries would show different characteristics from the ones described for DNA packaging machines. Although no single-molecule experiments have been reported up to date, the atomic resolution of two different RNA translocase conformations (P4 protein, see Sect. 12.4.1), and accurate biochemical studies have given strong structural insights into the molecular mechanistics of the dsRNA motors [27]. The main outcome from these data shows several interesting similarities with the DNA packaging motors. Structural snapshots showed that the P-loop of P4 changes its conformation from up to down in the presence of two different ATP analogues suggesting that this movement could help to pump the RNA inside the prohead (see Sect. 12.4 and Fig. 12.4). It has been proposed that P4 would follow a sequential translocation mechanism. In this mechanism the coordination between different subunits would be essential. Communication between adjacent subunits could happen through “arginine fingers” commonly found in ATPase motors. In this model ATP hydrolysis will trigger a conformational change that place the arginine finger of the adjacent subunit into the ATPase active site starting the next ATP hydrolysis [36]. It is supposed that the ssRNA structure could be similar to the A-dsRNA form presenting 11 bases per turn. It has been suggested that this symmetry mismatch would be solved by the L1 loop that would act as a grommet correcting the RNA position [39].

12.5.3 dsDNA Packaging Models

As previously described in Sect. 12.4, dsDNA packaging motors are more complex than dsRNA packaging motors. Usually their packaging machinery is composed by the dodecameric connector that sits in only one vertex of the prohead, and the terminase complex composed by the viral ATPase, which transiently associates to the prohead during packaging. Despite large efforts invested in the characterization of these motors, their precise molecular mechanism of translocation remains unclear. The models proposed can be classified in two categories. (1) Connector-driven models, where conformational changes induced by ATP hydrolysis in the connector structure lead to DNA packaging. (2) ATPase-driven models, where the connector is considered as a passive actor in the translocation mechanism while conformational changes in the ATPase directly translocate DNA. It is important to point out that both models consider the terminase as the ATP-consuming element of the motor. The difference lies in the element that actively pushes the DNA inside the capsid. In the case of the connector-driven models the terminase regulates the connector activity so it can push the DNA inside the capsid [50].

The first proposed model was based on the symmetry mismatch between the connector and the prohead, 12-fold and 5-fold respectively. This lack of symmetry would lead to a weak interaction between both components allowing the rotation of the connector inside the capsid. In this model connector rotation would lead to nucleic acid packaging inside the viral head [38]. The atomic structure determination of the connector allowed proposing later on the precise conformational changes inside the structure that would lead to DNA translocation. Observation of the Φ29 atomic structure showed that the helices that built the channel wall are not in a straight position. It was thus suggested that the straightening of one single helix by 12° rotation at the bottom of the connector would lead the expansion of the structure. This conformational change would be followed by the top of the protein leading to the contraction of the complex. This conformational change would be induced by the sequential hydrolysis of the ATP by the terminase (Fig. 12.8a) [38]. Another model based on the Φ29 connector structure suggested that the positive lysine ring inside the structure acts as an electrostatic grip on the negatively charged DNA molecule (Fig. 12.8b). This electrostatic interaction, together with the rotation of the connector would let the DNA to switch between two lysine rings and to be translocated without any additional connector structural rearrangement [50]. SPP1 connector pseudoatomic structure also inspired a different model in which the central channel is closed by a loop belt, suggesting that a sequential conformational change in these loops could act as a molecular lever to translocate DNA inside the capsid. In this model ATP hydrolysis will lead to 12° connector rotation and loop sliding and returning to the initial conformation [38].

Connector-driven models have been challenged after it was proved that the connector does not rotate inside the capsid. The recently solved structures of the T4 terminase have reinforced the terminase-driven models. The carefully biochemical and structural analysis carried out in T4 suggested the localization of the

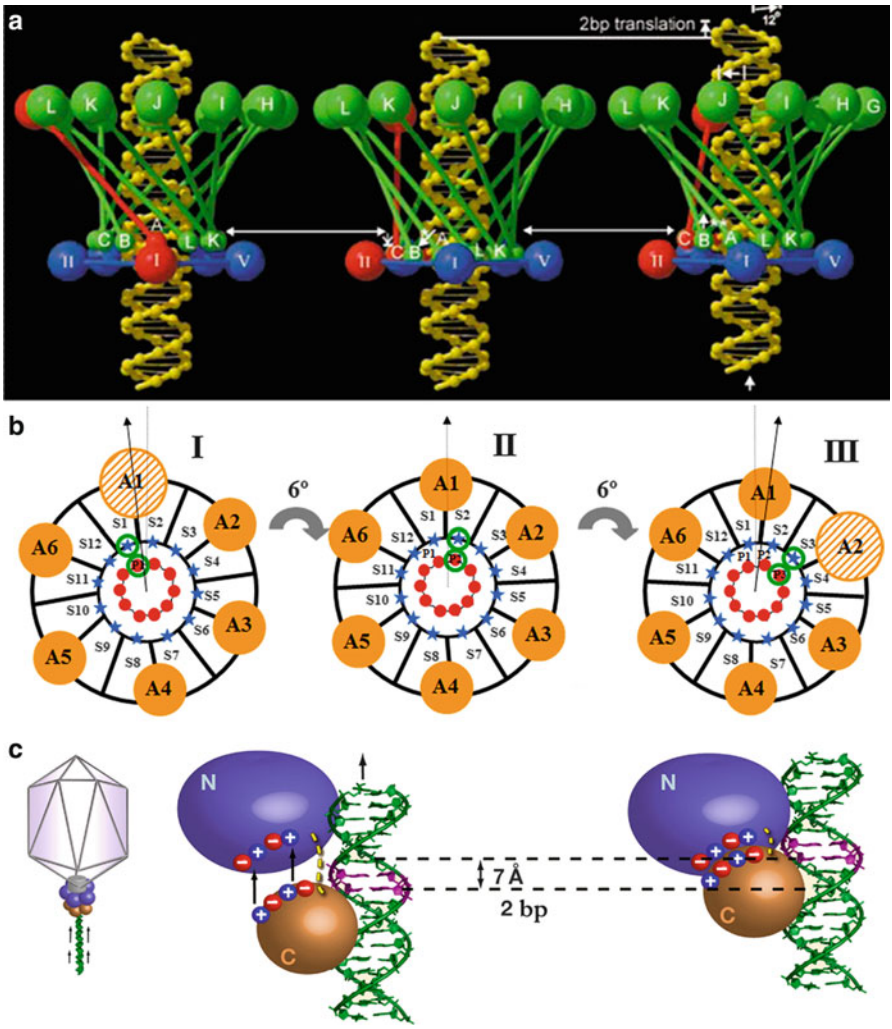


Fig. 12.8 Mechanical models of DNA packaging. (**a**) Compression-relaxation model showing the ATPase represented in *blue spheres* and the connector in *green*. The subunit of the connector suffering the conformational change is represented in *red*. First the subunit is straightened from the bottom (*centre*), then the movement is followed by the upper part consisting in translocation of 2 bp. (**b**) DNA gripping model showing the 12 subunits of the connector in *black* and the ATPase subunits as *yellow circles*. The ring of lysines located in the connector channel acts as an electrostatic docking point for the DNA (Figures in panels (a) and (b) are reproduced from [50]. With permission). (**c**) Terminase electrostatic model showing the two terminase domains Ndom in *blue* and Cdom in *orange*. During DNA packaging the two domains get closer and are pulled out by electrostatic repulsion allowing DNA translocation (Reproduced from [57]. With permission)

structural elements involved in ATP hydrolysis and their coupling to DNA translocation (see Sect. 12.4). It was shown that the terminase is composed by two domains: the Ndom which presents the ATPase activity and the nuclease Cdom (see Sect. 12.4). Both domains are joined by a flexible linker that allows their motility. It was proposed a model based on electrostatic interactions where ATP hydrolysis triggers a conformational change in the Ndom aligning both domains. This domain neighbouring will cause the pulling of the Cdom in an inch-worm like movement driven by the electrostatic forces conducing to DNA packaging (Fig. 12.8c) [58].

12.5.4 Differences Between Nucleic Acid Packaging and Ejection

Viral capsids are not impermeable containers and they present small pores that are necessary to exchange salts and ions with the media. Nevertheless these pores are too small to allow viral genome exit during infection. The infection process requires either capsid disassembly (as it occurs in most animal viruses) or the presence of a channel large enough to accommodate the viral genome. In viruses following an active packaging mechanism (dsDNA and dsRNA bacteriophages) the nucleic acid exits the capsid through the portal/translocase channel, at the fivefold vertex (see Chap. 17), the same passage that is used during genome packaging (see Sect. 12.4). These viruses package their genome in ordered layers (see Sect. 12.1) inside the capsid condensing its nucleic acid at a high density and accumulating an important potential energy (reflected in an internal pressure of up to 60 atmospheres; see Chaps. 9, 18, and 19). This high pressure requires the presence of protein complexes in the portal axis able to retain the nucleic acid inside the capsid. In dsDNA viruses this task is accomplished either by conformational changes in the portal channel or by protein plugs, like the tail proteins or the gatekeeper proteins [59] (see Chap. 17).

Even though nucleic acid packaging and ejection follow the same channel pathway they are not inverse processes. While packaging requires ATP consumption, genome delivery uses the energy stored in the nucleic acid during the condensation process [60]. Nevertheless this internal energy decreases as the genome is liberated and, thus, it is believed that other mechanisms may participate, as it is the case of proteins inside the host cell that pull the nucleic acid when the replication process is initiated (see [61] and Chap. 19). Differences are also evidenced by the fact that both processes have very different speeds, at least as measured in *in vitro* experiments: while DNA packaging is carried out at an average speed of 700 bp/s, DNA release can attain 60,000–70,000 bp/s.

In dsDNA bacteriophages there are strong evidences that the DNA entry and exit are carried out by different protein complexes. While DNA translocation is done by the packaging motor, DNA delivery uses the tail machine [62] (see Chap. 17). As some protein loops may as well be retaining the exit of the DNA, genome delivery requires conformational changes to modify the ejection channel in order to allow DNA passage [59]. Some viruses with dsRNA genomes present similar

characteristics, as in the first stages of the viral cycle they extrude their mRNAs trough the channels at the fivefold vertex (Fig. 12.2 and Sect. 12.3). During this process the positive strand from dsRNA is displaced to generate the mRNA substrate. Once more, this process is not energy-consuming and P4 protein serves as a passive channel for ssRNA.

12.6 Perspectives and Conclusions

Viruses protect their genetic material against the outside media or the cellular degrading enzymes by enclosing it inside a protein capsid which also serves other biological functions. Viral particles have evolved different packaging strategies according to their lifestyles, type of nucleic acid (RNA or DNA, single- or double-stranded), amount of genetic material to be packaged and other variables. A widespread strategy consists in the co-assembly of capsid and nucleic acid; in these cases the nucleic acid may act as a scaffold to promote capsid assembly, and the capsid proteins may help in the structuration of the nucleic acid inside the virion. In the simplest cases, this co-assembly strategy results in a helical structure that specifically encloses the viral nucleic acid. Packaging pathways involving condensation of capsid and nucleic acid are used by many bacterial, plant or animal ssRNA or ssDNA viruses.

Many other icosahedral viruses use dsRNA or dsDNA as genetic material. Packaging of long, stiff double-helical nucleic acid molecules pose additional problems; these viruses evolved sophisticated mechanisms for the packaging of these nucleic acids into preformed capsids. Viral particles actively packaging their double-helical nucleic acid genomes into spherical containers usually organize the nucleic acid in concentric layers inside the capsid. This arrangement facilitates genome uncoating by ejection without capsid disassembly.

The packaging machinery in those dsDNA and dsRNA viruses is a molecular motor that transforms the chemical energy of NTP hydrolysis into mechanical work to pump the nucleic acid inside the capsid. During the packaging process the initial immature capsid (prohead) suffers a series of conformational changes to become a mature capsid. The nucleic acid is transported inside the capsid through a channel formed by a protein complex present at a fivefold axes. These proteins present similarities with other nucleic acid translocases, showing ring-like structures with a central channel that serves for genome transport. In dsRNA bacteriophages this protein is named RNA-translocase and constitutes the simplest packaging motor described up to date. Structural snapshots obtained by X-ray crystallography suggest a mechanical mechanism consisting in the movement of a loop that would sequentially change conformation to transport the genome. dsDNA bacteriophages present a more complicated machinery, formed by a multi-protein complex that assembles transiently to the capsid. In these viruses the translocation pore is built by a protein, the connector, that sits at one single vertex of the capsid. The connector serves as a docking point for a viral complex, the terminase, that attaches transiently

to the capsid during packaging. The terminase complex is essentially composed by an ATPase and a secondary component, either a protein or RNA, which enhances the activity of the ATPase and specifically recognizes the viral genome. The properties of this multicomponent motor have been characterized using single molecule techniques (optical tweezers), which have revealed the biophysical characteristics of what constitutes one of the most powerful motors described to date. Several models have been proposed to couple the chemical energy of ATP hydrolysis with the mechanical action of DNA pumping; nevertheless the precise molecular mechanism remains unknown. The current hypothesis suggests that the connector will play a passive role during packaging helping with ATPase regulation, and in some cases acting as a sensor to indicate the degree of capsid filling. The active pumping of the DNA will be carried out by the ATPase structural reorganization moving the DNA inside the capsid.

Together, these results show that genome packaging is a very complex mechanism essential for virus survival. Its importance have forced some viruses to incorporate sophisticated machineries that share structural and mechanistic properties with other protein complexes playing essential tasks during the metabolism of genetic material in living cells.

Acknowledgements We acknowledge Jaime Martin-Benito for his contribution to Fig. 12.1c. This work was supported by Grant BFU2011-29038 from the Spanish Ministry of Science and Innovation.

References and Further Reading

1. Hunter E (2007) Virus assembly. In: Knipe DN, Howley PM (eds) Fields virology, vol 1. Lippincott Williams, Philadelphia, pp 141–168
2. Bhyravbhatla B, Watowich SJ, Caspar DL (1998) Refined atomic model of the four-layer aggregate of the tobacco mosaic virus coat protein at 2.4-Å resolution. Biophys J 74:604–615
3. Sachse C, Chen JZ, Coureux PD, Stroupe ME, Fandrich M, Grigorieff N (2007) High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. J Mol Biol 371:812–835
4. Palese P, Shaw ML (2007) Orthomyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields virology, vol 2. Lippincott Williams and Wilkins, Philadelphia, pp 1647–1689
5. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. Annu Rev Microbiol 60:51–67
6. Spencer SM, Sgro JY, Dryden KA, Baker TS, Nibert ML (1997) IRIS explorer software for radial-depth cueing reovirus particles and other macromolecular structures determined by cryoelectron microscopy and image reconstruction. J Struct Biol 120:11–21
7. Agirrezabala X, Martin-Benito J, Caston JR, Miranda R, Valpuesta JM, Carrascosa JL (2005) Maturation of phage T7 involves structural modification of both shell and inner core components. EMBO J 24:3820–3829
8. Lamb RA (2007) Mononegavirales. In: Knipe DM, Howley P (eds) Fields virology, vol 1. Lipincott Williams and Wilkins, Philadelphia, pp 1357–1361
9. Green TJ, Zhang X, Wertz GW, Luo M (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science 313:357–360

10. Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortin J, Martin-Benito J (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog 5: e1000491
11. Bunka DH, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, Twarock R, Phillips SE, Stockley PG (2011) Degenerate RNA packaging signals in the genome of satellite tobacco necrosis virus: implications for the assembly of a T=1 capsid. J Mol Biol 413:51–65
12. Dykeman EC, Grayson NE, Toropova K, Ranson NA, Stockley PG, Twarock R (2011) Simple rules for efficient assembly predict the layout of a packaged viral RNA. J Mol Biol 408:399–407
13. Prasad BV, Prevelige PE Jr (2003) Viral genome organization. Adv Protein Chem 64:219–258
14. Coombs KM (2006) Reovirus structure and morphogenesis. Curr Top Microbiol Immunol 309:117–167
15. Reinisch KM, Nibert ML, Harrison SC (2000) Structure of the reovirus core at 3.6 a resolution. Nature 404:960–967
16. Caston JR, Ghabrial SA, Jiang D, Rivas G, Alfonso C, Roca R, Luque D, Carrascosa JL (2003) Three-dimensional structure of penicillium chrysogenum virus: a double-stranded RNA virus with a genuine T=1 capsid. J Mol Biol 331:417–431
17. Sidhu SS (2001) Engineering M13 for phage display. Biomol Eng 18:57–63
18. Liu DJ, Day LA (1994) Pf1 Virus structure: helical coat protein and DNA with paraxial phosphates. Science 265:671–674
19. Chapman MS, Rossmann MG (1995) Single-stranded DNA-protein interactions in canine parvovirus. Structure 3:151–162
20. Johnson JE, Chiu W (2007) DNA packaging and delivery machines in tailed bacteriophages. Curr Opin Struct Biol 17:237–243
21. Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC (1997) Encapsidated conformation of bacteriophage T7 DNA. Cell 91:271–280
22. Petrov AS, Boz MB, Harvey SC (2007) The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. J Struct Biol 160:241–248
23. Comolli LR, Spakowitz AJ, Siegerist CE, Jardine PJ, Grimes S, Anderson DL, Bustamante C, Downing KH (2008) Three-dimensional architecture of the bacteriophage phi29 packaged genome and elucidation of its packaging process. Virology 371:267–277
24. Fang PA, Wright ET, Weintraub ST, Hakala K, Wu W, Serwer P, Jiang W (2008) Visualization of bacteriophage T3 capsids with DNA incompletely packaged *in vivo*. J Mol Biol 384:1384–1399
25. Poranen MM, Bamford DH (2012) Assembly of large icosahedral double-stranded RNA viruses. Adv Exp Med Biol 726:379–402
26. Huiskonen JT, de Haas F, Bubeck D, Bamford DH, Fuller SD, Butcher SJ (2006) Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. Structure 14:1039–1048
27. Kainov DE, Tuma R, Mancini EJ (2006) Hexameric molecular motors: P4 packaging ATPase unravels the mechanism. Cell Mol Life Sci 63:1095–1105
28. Mindich L (2012) Packaging in dsRNA viruses. Adv Exp Med Biol 726:601–608
29. Ionel A, Velazquez-Muriel JA, Luque D, Cuervo A, Caston JR, Valpuesta JM, Martin-Benito J, Carrascosa JL (2011) Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7. J Biol Chem 286:234–242
30. Valpuesta JM, Carrascosa JL (1994) Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Quart Rev Biophys 27:107–155
31. Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF (2005) Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr Opin Struct Biol 15:227–236
32. Gertsman I, Gan L, Guttman M, Lee K, Speir JA, Duda RL, Hendrix RW, Komives EA, Johnson JE (2009) An unexpected twist in viral capsid maturation. Nature 458:646–650

33. Lawton JA, Estes MK, Prasad BV (2000) Mechanism of genome transcription in segmented dsRNA viruses. Adv Virus Res 55:185–229
34. Mancini EJ, Kainov DE, Grimes JM, Tuma R, Bamford DH, Stuart DI (2004) Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. Cell 118:743–755
35. McDonald SM, Patton JT (2011) Assortment and packaging of the segmented rotavirus genome. Trends Microbiol 19:136–144
36. Mancini EJ, Tuma R (2012) Mechanism of RNA packaging motor. Adv Exp Med Biol 726:609–629
37. Cuervo A, Carrascosa JL (2012) Bacteriophages: structure. In: eLS. Wiley, Chichester, pp 1–7
38. Rao VB, Feiss M (2008) The bacteriophage DNA packaging motor. Annu Rev Genet 42:647–681
39. Sun S, Rao VB, Rossmann MG (2010) Genome packaging in viruses. Curr Opin Struct Biol 20:114–120
40. Mettenleiter TC, Klupp BG, Granzow H (2006) Herpesvirus assembly: a tale of two membranes. Curr Opin Microbiol 9:423–429
41. Cuervo A, Carrascosa JL (2012) Viral connectors for DNA encapsulation. Curr Opin Biotechnol 23:529–536
42. Carrascosa JL, Valpuesta JM (1999) Bacteriophage connectors: structural features of a DNA translocating motors. In: recent research developments in virology. Transworld reseach network. Trivadrum 1:449–465
43. Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. Adv Exp Med Biol 726:489–509
44. Ding F, Lu C, Zhao W, Rajashankar KR, Anderson DL, Jardine PJ, Grimes S, Ke A (2011) Structure and assembly of the essential RNA ring component of a viral DNA packaging motor. Proc Natl Acad Sci U S A 108:7357–7362
45. Casjens SR (2011) The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol 9:647–657
46. Rao VB, Black LW (2010) Structure and assembly of bacteriophage T4 head. Virol J 7:356
47. Nemecek D, Lander GC, Johnson JE, Casjens SR, Thomas GJ Jr (2008) Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. J Mol Biol 383:494–501
48. Buttner CR, Chechik M, Ortiz-Lombardia M, Smits C, Ebong IO, Chechik V, Jeschke G, Dykeman E, Benini S, Robinson CV, Alonso JC, Antson AA (2012) Structural basis for DNA recognition and loading into a viral packaging motor. Proc Natl Acad Sci U S A 109:811–816
49. Anderson D, Grimes S (2005) In: Catalano CE (ed) Viral Genome packaging machines: genetics, structure and mechanism. Kluwer Academic/Plenum Publishers, New York
50. Morais MC (2012) The dsDNA packaging motor in bacteriophage ϕ29. Adv Exp Med Biol 726:511–547
51. Guo P, Lee TJ (2007) Viral nanomotors for packaging of dsDNA and dsRNA. Mol Microbiol 64:886–903
52. Shu D, Zhang H, Jin J, Guo P (2007) Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. EMBO J 26:527–537
53. Hingorani MM, O'Donnell M (1998) Toroidal proteins: running rings around DNA. Curr Biol 8:R83–R86
54. Massey TH, Mercogliano CP, Yates J, Sherratt DJ, Lowe J (2006) Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. Mol Cell 23:457–469
55. Gomis-Ruth FX, Moncalian G, Perez-Luque R, Gonzalez A, Cabezon E, de la Cruz F, Coll M (2001) The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. Nature 409:637–641
56. Chemla YR, Smith DE (2012) Single-molecule studies of viral DNA packaging. Adv Exp Med Biol 726:549–584
57. Williams RS, Williams GJ, Tainer JA (2008) A charged performance by gp17 in viral packaging. Cell 135:1169–1171

58. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB (2008) The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. Cell 135:1251–1262
59. Tavares P, Zinn-Justin S, Orlova EV (2012) Genome gating in tailed bacteriophage capsids. Adv Exp Med Biol 726:585–600
60. Purohit PK, Inamdar MM, Grayson PD, Squires TM, Kondev J, Phillips R (2005) Forces during bacteriophage DNA packaging and ejection. Biophys J 88:851–866
61. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJ (2007) Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. Cell Mol Life Sci 64:1484–1497
62. Vinga I, Sao-José C, Tavares P, Santos M (2006) Bacteriophage entry in the host cell. In: Wegrzyn G (ed) Modern bacteriophage biology and biotechnology. Research Signpost, Kerala, pp 165–205

Further Reading

Flint SJ, Enquist LW, Racaniello VR, Skalka AM (eds) (2009) Principles of virology, 3rd edn. ASM Press, Washington
Knipe DM, Howley PM (eds) (2007) Fields virology, 5th edn. Lippincott Williams & Wilkins, Philadelphia
Patton JT (ed) (2008) Segmented double-stranded RNA viruses. Structure and molecular biology. Caister Academic Press, Norfolk
Rossmann MG, Rao VB (eds) (2012) Viral Molecular Machines. Adv Exp Med Biol vol. 726, Springer, New York

Also especially recommended for further reading are references [5, 20, 27, 45] *listed above*.

Chapter 13
Virus Maturation

Laura R. Delgui and José F. Rodríguez

Abstract The formation of infectious virus particles is a highly complex process involving a series of sophisticated molecular events. In most cases, the assembly of virus structural elements results in the formation of immature virus particles unable to initiate a productive infection. Accordingly, for most viruses the final stage of the assembly pathway entails a set of structural transitions and/or biochemical modifications that transform inert precursor particles into fully infectious agents. In this chapter, we review the most relevant maturation mechanisms involved in the generation of infectious virions for a wide variety of viruses.

Keywords Capsid • Capsomer • Envelope • Glycoprotein • Occlusion body • Polyhedra • Polyprotein • Procapsid • Protease • Scaffold • Structural rearrangement • Virion • Virus egress • Virus entry • Virus maturation • Virus morphogenesis.

Abbreviations

3D	Three dimensional.
AcMNPV	*Autographa californica* nucleopolyhedrovirus
CA	Capsid protein of HIV
DNA	Deoxyribonucleic acid
ER	Endoplasmic reticulum

L.R. Delgui
Laboratorio de Biología Celular y Molecular, Instituto de Histología y Embriología Mendoza (IHEM), Facultad de Ciencias Médicas, Instituto de Ciencias Básicas, Universidad Nacional de Cuyo, Mendoza, Argentina

J.F. Rodríguez (✉)
Department of Molecular and Cell Biology, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jfrodrig@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_13,

ESCRT	Endosomal sorting complexes required for transport
gp	Glycoprotein
HA	Hemagglutinin
HB-sAg	HBV surface antigen
HBV	Hepatitis B virus
HIV-1	Human immunodeficiency virus type 1
HSV-1	Herpes simplex virus type 1
IAV	Influenza A virus
kbp	Kilobase pairs
MA	Matrix protein of HIV
NC	Nucleocapsid protein of HIV
NωV	*Nudaurelia capensis* ω virus
ORF	Open reading frame
PR	Protease of HIV
RNA	Ribonucleic acid
T	Triangulation number
VLP	Virus-like particle
VP	Virus protein
WNPV	*Wiseana* nucleopolyhedrovirus

13.1 Introduction

The requirement of a maturation step during virus morphogenesis is directly related with the building strategy used by most viruses. As for many other large multi-subunit complexes, the assembly of virus particles is initiated by the establishment of weak interactions between structural subunits. This is critical for the correct positioning of the particle building blocks and the prevention of the formation of aberrant structures during the assembly process [39] (see Chaps. 10 and 11, 19). As a result of this initial step, labile immature particles, generally known as procapsids or provirions, are formed (Chaps. 10 and 11).

Due to their intrinsic fragile nature, procapsids are unable to withstand neither the internal pressure that may be exerted in those cases where the viral genome is tightly packaged (see Chap. 12) nor the external physical challenges inherent to the virus life-cycle that, in most cases, involve: (i) the egress of the particles from the infected cell to the extracellular milieu; (ii) survival in a highly hostile external environment; and (iii) interaction(s) with cellular receptors and the subsequent entry into a new host cell (Chaps. 15–17). The maturation process, generally triggered by one or more proteases present within the procapsid, involves a solid-state phase transition that, depending on the virus model, varies from subtle to major conformational rearrangements. This process takes place in a highly

ordered stepwise manner designed to consolidate the virion structure, and thus allowing it to achieve the structural properties required to be fully infectious [42].

The maturation of enveloped viruses generally requires a further step. To become infectious, glycoprotein envelope components involved in cell receptor interactions must undergo complex glycosylation pathways for their egress from the host cell. Additionally, these viruses depend upon proteolytic processing of their glycoprotein components for the activation of the membrane fusion events required for the internalization of the nucleocapsid to cell cytoplasm [43] (see Chap. 16).

Finally, some viruses secure their extracellular endurance by including a fraction of the newly formed virus progeny into discrete, highly resistant virus-derived proteinaceous superstructures. These superstructures, generally known as polyhedra or occlusion bodies, play a major role in the virus life-cycle providing a long-term stability that allows occluded virions to survive harsh environmental conditions for prolonged periods of time [8].

13.2 Immature Virus Particles and Maturation Strategies

Mechanisms directing the assembly of immature particles largely depend upon the structural complexity of the virus particle. Thus, whilst procapsids from simple viruses such as parvoviruses (see Chap. 10), nodaviruses or tetraviruses are capable of self-assembling without a direct contribution of auxiliary elements, the assembly of more complex viruses requires the participation of virus-derived structural components, *i.e*. the virus genome (see Chap. 12) and/or scaffolding polypeptides (see Chap. 11), different from those found as integral constituents in the mature capsid. As described above, maturation is a transition process that transforms inert immature particles into fully infectious virions. Indeed, regardless of their structural complexity, maturation processes have been documented in all virus models for which the assembly pathway has been characterized in depth.

The maturation of procapsids from simple viruses generally involves the autocatalytic processing of capsid polypeptides. This increases the stability of the particle and confers full virus infectivity. The presence of scaffolding proteins in procapsids from more complex viruses is essential to ensure the fidelity of the assembly process, and thus avoid the generation of aberrant dead-end structures. A comprehensive review about the role and structure of scaffolding polypeptides has been recently published [24]. Scaffolding polypeptides can be categorized in two main groups: (i) icosahedrally ordered, external scaffolds; and (ii) internal, core-like, scaffolds. The highly complex prolate cores found in T4-like bacteriophages also playing a critical scaffolding role and not fitting the categories described above have been described in detail in Chap. 11. The function of scaffolding polypeptides is intrinsically transient. Hence, during maturation they are either expelled out from the procapsid or displaced to a different topological location within the structure of the mature virion. Although the maturation of some viruses does not involve the

cleavage of scaffolding elements, in most cases it requires the activation of one or more well-coordinated sets of proteolytic processing events. This brings about the alteration of both particle size, either expansion or contraction, and of the shape of the original procapsid structure. Finally, maturation enveloped viruses involves an additional step consisting on the proteolytic cleavage of glycosylated envelope polypeptides responsible for the interaction with specific cell receptors, and the subsequent fusion of the virus envelope and the host cell membrane.

Following there is a description of well-characterized maturation processes corresponding to three virus groups with increasing levels of complexity.

13.3 Tetravirus Capsid Maturation

Tetraviruses are small positive-stranded RNA viruses that infect different insect species belonging to the *Lepidoptera* order [16]. Tetravirus virions are nonenveloped icosahedrons with a T = 4 symmetry enclosing a bipartite positive-stranded RNA genome. The virus particle is built by 240 copies of the capsid polypeptide (70-kDa) known as the α protein. During particle assembly, the α protein undergoes an autocatalytic processing event, directly associated to particle maturation, that generates the β (62-kDa) and the γ (8-kDa) polypeptides [1, 31].

Characterization of the tetravirus maturation process has mainly used the *Nudaurelia capensis* ω virus (NωV) as a model. The NωV capsid assembly and maturation process has been finely dissected using recombinant baculovirus-based expression systems. Expression of the NωV coat polypeptide in insect cells from recombinant baculoviruses leads to the assembly of virus-like particles (VLP) that, when exposed to an acidic environment, mature to form particles structurally identical to authentic virions [44].

The NωV procapsid to capsid transition involves a major, pH-induced, conformational change involving large-scale movements that drastically reduce the particle diameter (>15 %), and modify its external appearance. VLPs purified at near-neutral pH (pH = 7.6) have a diameter of 485 Å, a highly spherical appearance and a conspicuously porous surface (Fig. 13.1). These VLPs are thought to faithfully mimic the procapsid transient structure produced during NωV assembly in infected cells, a process that takes place in an acidic cellular compartment. The capsid protein subunits that form baculovirus-derived procapsids remains in its original uncleaved α state. Incubation of procapsids in acidic solutions (pH = 5) causes a major structural transition. The resulting particles have a diameter of 395 Å, a well-defined icosahedral aspect, and a smooth surface (Fig. 13.1). Although the structural transition from procapsid to capsid is very fast, taking place within the first 100 milliseconds of incubation at low pH, the self-processing of the coat protein is a much slower process that requires several hours to be completed. Interestingly, the procapsid to capsid transition is reversible, and capsids can be re-expanded to the procapsid stage by raising the pH. The transition becomes irreversible only when ca. 15 % of the

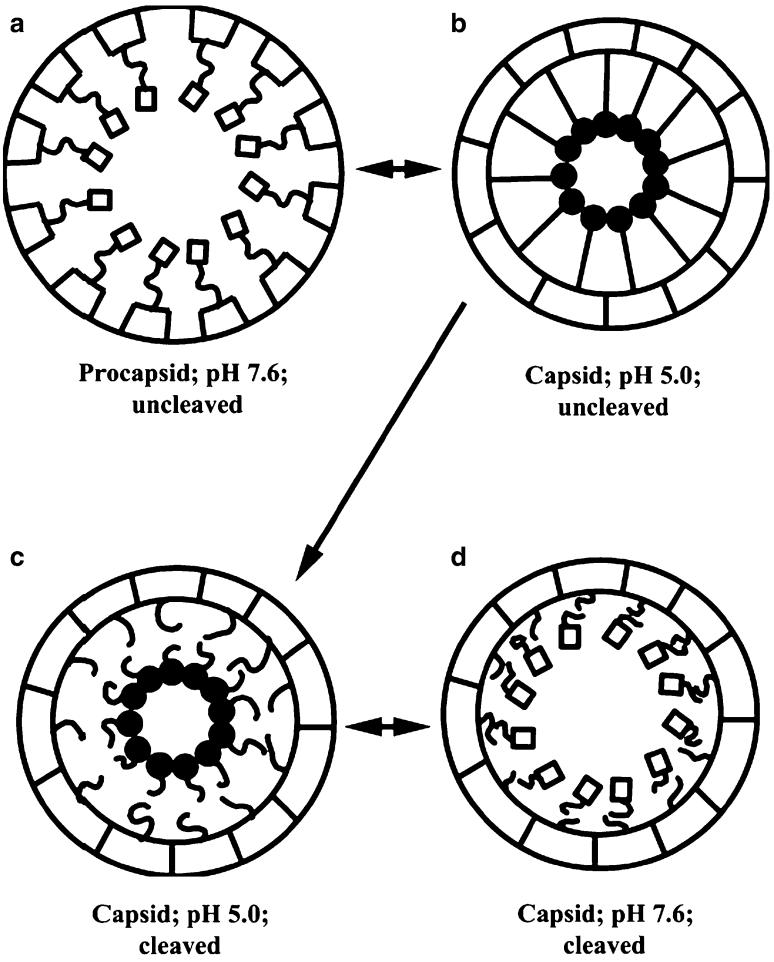


Fig. 13.1 Tetravirus capsid maturation. Schematic representation of the relationship of coat protein cleavage and the pH-induced conformational change in NωV. The region of the helical domain of each coat protein subunit believed to be responsible for the helix-coil transition represents residues 1–44 and 571–644 of the 644-residue protein. The diagram depicts this region as either α-helix (*opened squares*) or a random coil (*black circles*). (**a**) The particles were purified as procapsid at pH 7.6, with no cleavage occurring in the subunits. (**b**) Upon lowering the pH to 5.0, protonation of specific residues causes the helix to transform into a coil, initiating the quaternary structural rearrangement. If no cleavage occurs in the capsid at pH 5.0, the process is reversible. (**c**) Cleavage of wild-type particles locks the particles in the capsid conformation. (**d**) Altering the pH of the particles allows the helix-coil to operate reversibly, but cleavage uncouples this pH-driven engine from the rest of the cargo subunit, causing the surface of the particles to resemble capsids (Reproduced from [44]. With permission)

procapsid α coat protein subunits have been proteolytically cleaved and converted into the mature capsid β protein form.

Expression of mutant versions of the α coat protein lacking the self-cleavage site leads to the assembly of procapsids structurally identical to those obtained by expression of the wild-type α polypeptide. These mutant procapsids also undergo the procapsid to capsid transition when incubated at low pH. However, the lack of the scissile bond for the α protein autoproteolytic processing completely blocks the consolidation of the mature capsid stage [28].

Comparisons of procapsid and capsid secondary structures using Fourier transform infrared spectroscopy revealed that the α-helical content of the procapsid is higher than that of the capsid, thus indicating that local refolding of internal helical regions of the α capsid protein to coil is critical for the maturation process. Two alternative models have been proposed to explain the maturation process. The first one assumes that a helix to coil transition of γ-peptide region of the immature α capsid proteins found in the procapsid constitutes the driving force for the capsid rearrangement [44]. In this scenario, the autoproteolytic cleavage of the α protein would switch off the driving force for conformational change from the rest of the capsid, thus making the transition irreversible once cleavage has occurred. An alternative model, based on information gathered from comparisons of the crystal structure of both procapsids and capsids, suggests that charge repulsion amongst clusters of acidic amino acid residues found at subunit interfaces are responsible for particle re-expansion when incubated above their isoelectric point [19].

A very important aspect of the NωV capsid maturation process is the observation that although mature capsid particles are more stable to pH and ionic conditions as well as more resistant to proteolysis than procapsids, they are far more sensitive than the latter to thermal stress. This suggests that the end product of the maturation process is the formation of a metastable virion capable of safely transporting the virus genome to its target cell but designed to easily release its genetic cargo upon the correct physical stimuli, *i.e.* the interaction with its cognate receptor(s).

13.4 Herpesvirus Nucleocapsid Maturation

Members of the *Herpesviridae* family are highly complex double-stranded DNA viruses infecting a wide variety of animal species [11]. Viruses from the three herpesvirus subfamilies - *Alpha*-, *Beta*-, and *Gammaherpesvirinae*- exhibit significant genome conservation, and a common virion morphology. The innermost virion structure is the icosahedral capsid that contains the viral DNA and is surrounded by the tegument, a proteinaceous matrix lacking a defined structure, and by a lipidic envelope.

The herpesvirus nucleocapsid assembly process, that strongly resembles that of double-stranded DNA bacteriophages [35], has been characterized in detail using the human herpes simplex virus type 1 (HSV-1), prototype of the *Simplexvirus* genus of the *Alphaherpesvirinae* family, as virus model. The mature HSV-1 capsid

is an unusually large and complex structure of 125 nm in diameter, with a T = 16 icosahedral surface lattice. The capsid is built by two types of subunits: (i) hexamers and pentamers formed by VP5, the major capsid polypeptide; and (ii) triplexes, located between and connecting pentons and hexons, formed by a heterotrimer built by one VP19c molecule and a two copies of the VP23 polypeptide (Fig. 13.2a). One of the 12 particle vertices is occupied by the portal complex, a dodecameric structure that forms a channel that is essential for the packaging, and probably the release, of the viral DNA genome [4].

HSV-1 capsid assembly takes place in the nucleus of infected cells and is initiated by formation of a spherical procapsid containing two concentric protein layers: the procapsid shell and an underlying scaffold. In addition to proteins found in the mature capsid, procapsid assembly involves the participation of two additional virus-encoded polypeptides, pre-VP22a, the scaffolding polypeptide, and pre-VP21, the precursor of the protease, responsible for procapsid maturation. Although, as shown by experiments carried out using baculovirus-based procapsid artificial assembly systems [46], the presence of the portal complex is not essential for procapsid assembly. However, under physiological conditions it appears to play an important role for the initiation of the assembly process. Indeed, the portal complex is absolutely indispensable for the incorporation of the 152-kbp virus genome into the nascent procapsid, and for the subsequent cleavage and extrusion of the scaffolding polypeptide under native conditions.

The procapsid to capsid maturation process involves a massive, cooperative and irreversible rearrangement of the capsid shell. This process is promoted by the activation of the protease that releases the interaction between the scaffold and the shell layers, thus triggering a series of conformational changes, mainly relative rotations, affecting the major capsid protein VP5 forming the capsomers.

It has been shown that procasids produced by a protease-defective temperature-sensitive HSV-1 mutant can be purified from infected cells grown at non-permissive temperature. These procapsids are able to slowly mature when incubated at permissive temperature, thus allowing a close scrutiny of the maturation process using cryo-electron microscopy and three-dimensional (3D) image reconstruction [14, 21] (see Chap. 3). Comparison of the 3D images allowed assessing the existence of 17 clearly distinguishable intermediate maturation stages. In the procapsid, the neighboring capsomers (hexamers and pentamers) are held together mainly by interactions with the surrounding triplexes that exhibit a clear morphogenetic role. Upon maturation, triplexes are transformed into molecular clamps stabilizing the interaction of their three surrounding capsomers that come together at the inner base of the capsid layer. Indeed, the most conspicuous change associated to maturation is the formation of a floor, a continuous network of VP5 interactions, at the base of the capsid layer that reinforces particle stability. To form the floor connections, the VP5 base domains must rotate ca. 40° with respect to their external domains. A second major rotation affects the protrusion domains of the VP5 molecules found at hexons. In addition to their increased stability, at its

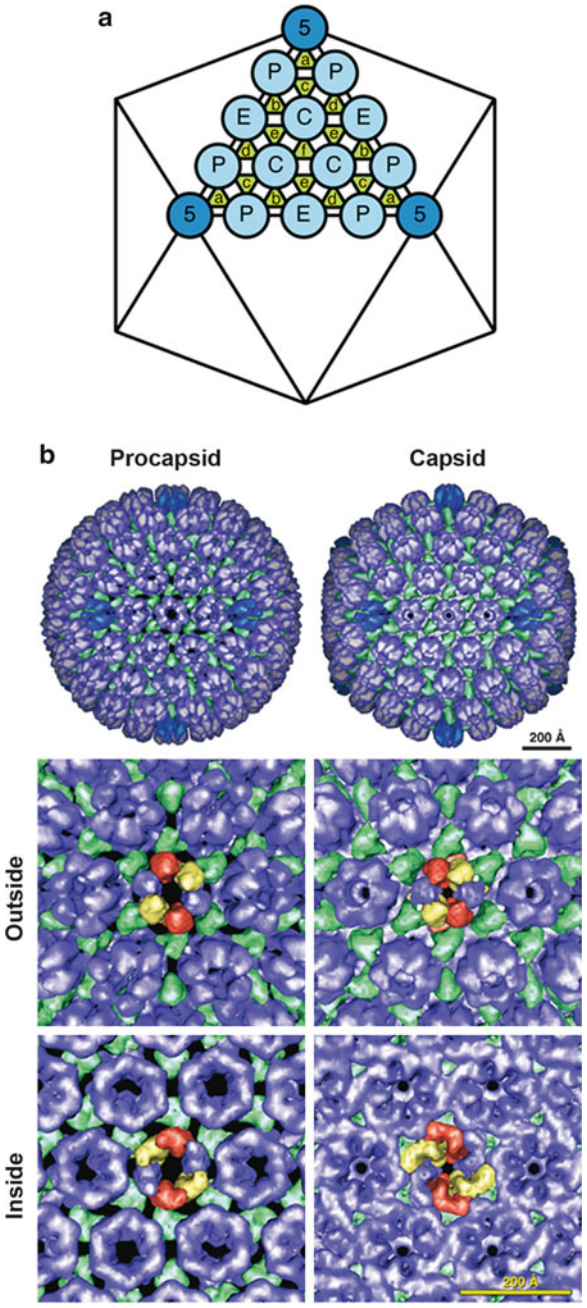


Fig. 13.2 Herpes simplex virus capsid maturation. (**a**) Herpes virus surface lattice. Diagram showing the positions of the quasi-equivalent capsomers and triplexes. The peripentagonal, edge and central hexons and pentons are denoted by P, E, C and 5, respectively. The six triplexes, Ta-Tf, are denoted by a–f (Reproduced with permission from [21]). (**b**) Molecular anatomy of the procapsid and the mature capsid. VP5 hexons are colored in *light blue*, pentons in *dark blue*, and triplexes in *green*. Particles are viewed along a two-fold symmetry axis. (**c**) Images show enlarged areas centered on an E hexon at a two-fold symmetry axis for the procapsid and the mature capsid and for the outer and inner surfaces, respectively. The three quasi-equivalent pair of VP5 subunits in the E hexon are *yellow*, *red* and *blue*, respectively. Other VP5 subunits are *blue* (Reproduced from [21]. With permission)

final maturation stage capsids have a more compact aspect, lacking intercapsomer gaps, and show an icosahedral contour that differentiates them from the original more globular procapsids (Fig. 13.2b).

13.5 Maturation of the Human Immunodeficiency Virus Capsid

The human immunodeficiency virus type 1 (HIV-1) is the best-characterized member of the Retrovirus family, a diverse group of enveloped viruses with a positive-stranded RNA genome. Retroviruses share a similar virion structure enclosing a dimeric genome, and some general replicative properties, *i.e.* the reverse transcription and the ability to integrate their genomes into the DNA of their host cells [45]. The highly complex HIV-1 assembly and maturation processes have been recently reviewed in depth [3].

The HIV-1 particle has a globular appearance with a mean diameter of about 120 nm. The envelope contains about 70 copies of the protruding Env protein complex which is built by a trimer of the glycoprotein gp120 bound to a trimer of the transmembrane gp41 polypeptide that forms a stem-like structure and anchors the complex to the viral envelope [5, 6]. The inner region of the viral particle is formed by proteins derived from the group-specific antigen (Gag) polypeptide, the structural polyprotein precursor. The geometric structure of the mature HIV-1 capsid is a fullerene cone, a conical hexagonal net closed at both ends through the introduction of 12 pentagonal defects [13]. This structure, built by ca. 1.500 molecules of the capsid protein (CA) in the form of hexameric and pentameric rings, encloses the molecular replicative machine, a ribonucleoprotein complex formed by nucleocapsid protein (NC) tightly bound to the virus genome and associated to the reverse transcriptase, the integrase, the protease [7] and the accessory protein Vpr. The core is surrounded by a discontinuous layer formed by molecules of matrix protein (MA).

The Gag protein precursor encompasses the above-mentioned proteins MA, CA and NC as three major structural components: MA, the membrane binding domain; CA, the capsid domain; and NC, the nucleocapsid domain, which interacts with and recruits the viral RNA. CA and NC are separated by a linker peptide, SP1, and downstream of NC are two further peptide domains, SP2 and p6 (Fig. 13.3a, b). p6 is responsible for the recruitment of the cellular endosomal sorting complexes required for transport (ESCRT) elements. The assembly of the spherical HIV-1 procapsid is exclusively driven by Gag precursor proteins, Gag (Pr55) and Gag–Pro-Pol (Pr160), targeted to the plasma membrane by myristoylation of their N-terminal domains. Accumulation of Gag precursors promotes the assembly of incomplete spheres underneath specialized host cell membrane microdomains containing virus-encoded glycoproteins. The assembly of these immature structures prompts the recruitment of components of the ESCRT machinery that drive

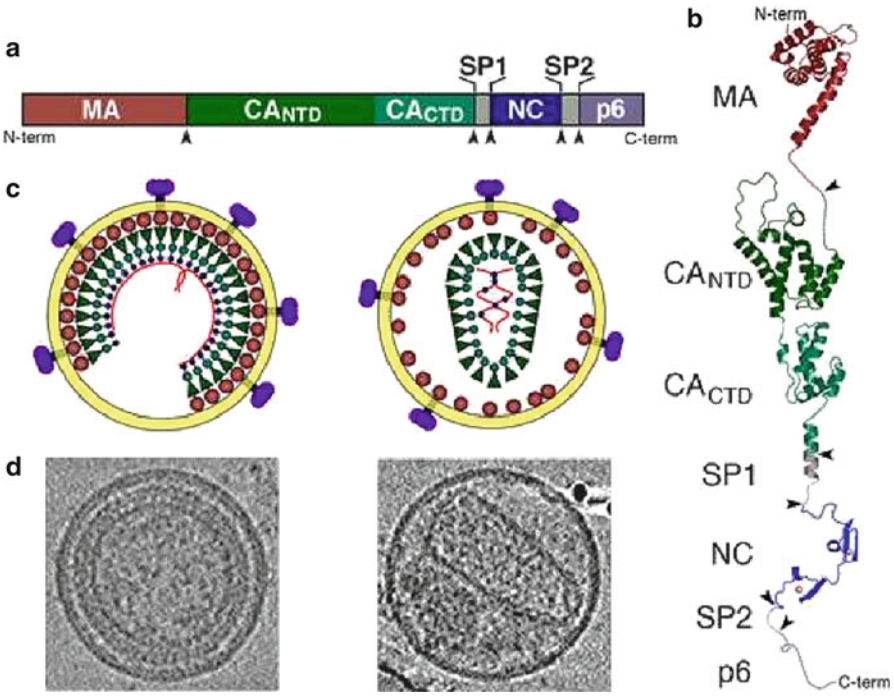


Fig. 13.3 HIV-1 particle maturation. (**a**) HIV-1 Gag polyprotein domain structure, showing the locations of MA, CA (N-terminal domain CANTD and C-terminal domain CACTD), SP1, NC, SP2, and p6. (**b**) Structural model of the extended Gag polypeptide, derived from high-resolution structures and models of isolated domains. *Dashed lines* represent unstructured and linker regions. PR cleavage sites are indicated by the *arrowheads* in (**a**) and (**b**). (**c**) Schematic models of the immature (*left*) and mature (*right*) HIV-1 virions. (**d**) Central slices through cryo-electron microscopy tomograms of immature (*left*) and mature (*right*) HIV-1 particles. The spherical virions are approximately 130 nm in diameter (Reproduced from [13]. With permission)

the scission of the membrane and the release of budding non-infectious immature virus particles from the cell [13] (see also Chap. 14). The Gag molecules in the immature particle are extended and oriented radially, with their N-terminal MA domains associated with the viral membrane and their C-terminal p6 domains facing the interior of the particle (Fig. 13.3c, d). The immature capsid lattice is stabilized primarily by lateral protein-protein interactions.

As it is the case for the maturation of procapsids of many icosahedral viruses, maturation of the HIV-1 procapsid involves a series of dramatic morphological changes and is directly associated to the activation of the protease (PR) activity that takes place during or immediately after budding. Although the PR polypeptide has been extensively characterized, its activation mechanism is not completely understood. In order to be fully active, PR has to dimerize and the activity in the Gag-Pro-Pol precursor is negligible. It is thought that Gag trafficking regulates PR activation by preventing premature PR dimerization until the Gag molecules coalesce at the

plasma membrane, and that an as yet unknown mechanism(s), other than the assembly-mediated PR dimerization, must play an important role in this process. The activated PR cleaves Gag in five positions, thus releasing the MA, CA, NC, and p6 proteins as well as the SP1 and SP2 peptides. Indeed proteolytic maturation is essential for infectivity, and PR inhibitors are of outmost importance in current antiretroviral therapies.

The five proteolytic cleavage sites in Gag are cleaved at very different rates *in vitro*. The fastest cleavage is that between SP1 and NC taking place 400-fold faster than that between CA and SP1 [34]. Additionally, the detection of reproducible processing intermediates in lysates from infected cells or upon partial inhibition of PR, suggests that cleavage is a stepwise process. The order of cleavage it is thought to involve the initial processing separating NC-p6 from the membrane-bound N-terminal part of Gag. The secondary cleavages would separate MA from CA-SP1 and p6 from NC-SP2, and final processing would release the two spacer peptides from the C-termini of CA and NC, respectively. After proteolytic cleavage, the MA layer is thought to remain associated with the viral membrane, whereas NC and the RNA are condensed into the ribonucleoprotein complex layer surrounded by the viral CA capsid core.

13.6 The Role of Glycosylation in Virus Maturation

One very important structural feature of many viruses is that their surface is covered with glycoproteins, and glycosylation is vital for virus replication and infectivity. Proper protein folding, attachment to host-cell receptors and evasion of host-immune responses are events where glycosylation is involved in some viral systems. In recent years, glycosylation has become an additional means of therapy by interfering with viral host cell entry or egress or by preventing the correct assembly of virions [30].

13.6.1 Hepadnavirus Glycosylation: Its Importance in Virus Assembly and Egress

Hepatitis B virus (HBV), a member of the *Hepadnaviridae* family, is a human pathogen causing acute and chronic liver disease that eventually leads to liver cirrhosis and hepatocellular carcinoma [12]. HBV particles are double-shelled spheres with a DNA-containing inner nucleocapsid and an outer envelope composed of cellular lipids and three structurally related virus-encoded proteins termed small (S), middle (M) and large (L) proteins, which are together referred as to as HBV surface antigen (HB-sAg). Virion assembly is initiated by insertion of the envelope proteins into the endoplasmic reticulum (ER) membrane proceeding at

pre-Golgi membranes, where cytosolic nucleocapsids are packaged by transmembrane envelope proteins. Virions then bud into intraluminal cisternae and leave the cell *via* the constitutive secretory pathway [23]. L, M and S proteins are derived from a single open reading frame (ORF) by employing three different translation start sites that divide the ORF into three domains: the N-terminal pre-S1 domain; the middle pre-S2 domain; and the C-terminal S domain common to S, M and L proteins (Fig. 13.4a) [29]. All three proteins possess a partially utilized *N*-glycosylation site at Asn-146 of the S domain and are also thought to have similar three-dimensional structures. The major difference between S and M proteins is the presence of 55 additional amino acid residues corresponding to the pre-S2 domain on M, and importantly the additional glycan site within this region at Asn-4. Although the L protein also contains the pre-S2 domain, this glycosylation site is not utilized (Fig. 13.4b) [18]. Under normal circumstances, after synthesis, HBsAg leaves the ER, passes through the Golgi stacks and is secreted within 3 h. However, when infected cells are treated with glucosidase inhibitors, enzymes that mediate the first steps in the trimming of terminal sugars pathway, the HBsAg molecules are detained and displaced to the Golgi, and even returned to the ER [25, 26]. On the other hand, detailed studies based on the use of site-directed mutagenesis revealed that the common glycan in L, M and S proteins does not play a role in virus secretion. Instead, the removal of the pre-S2 glycan site prevents the secretion of enveloped virus, rendering the M glycosylation at the ER as a crucial event in the formation of the HB viral particle [29].

For glycoproteins, the processing of the initial oligosaccharide precursor from the $Glc_3Man_9GlcNAc_2$ to the $Glc_1Man_9GlcNAc_2$ glycoform in the ER can lead to an interaction with chaperones such as calnexin. Calnexin binds only to glycoproteins containing $Glc_1Man_9GlcNAc_2$ structures to assist in their folding and anchors the polypeptides to the ER until they have achieved their correct folding conformation [17]. The HBV envelope M glycoprotein associates to calnexin. This interaction strictly depends on the glycan at Asparagine (Asn)-4, specific for M, while the common Asn-146-linked glycan is not involved [48]. It is hypothesized that the M protein acts as a dominant negative scaffolding glycoprotein, and that its misfolding destabilizes the viral envelope, thus hindering viral particle secretion.

A second glycosylation modification also restricted to the M protein has been described. It has been shown that the M proteins carry, at least in part, a single *O*-linked carbohydrate substituent, which could be identified as GalNAcα-, Gal(β1-3)GalNAcα- or Neu5Ac(α2-3)Gal(β1-3)GalNAcα-unit with the threonine in position 37 (Thr-37) of pre-S2 having the highest potential to be a O-glycosylated site [38]. Glycoproteins with O-linked glycans have been found in a number of enveloped viruses, although the distinct functions of viral *O*-glycoproteins remain obscure. For pre-S2 *O*-glycan, a masking of the respective peptide sequence has been suggested. Supporting this hypothesis mouse monoclonal antibodies directed against epitopes involving Thr-37 or neighboring amino acid residues have never been identified. In contrast, antibodies recognizing an epitope encompassing the *N*-glycan linked to Asn-4 are readily available [38].

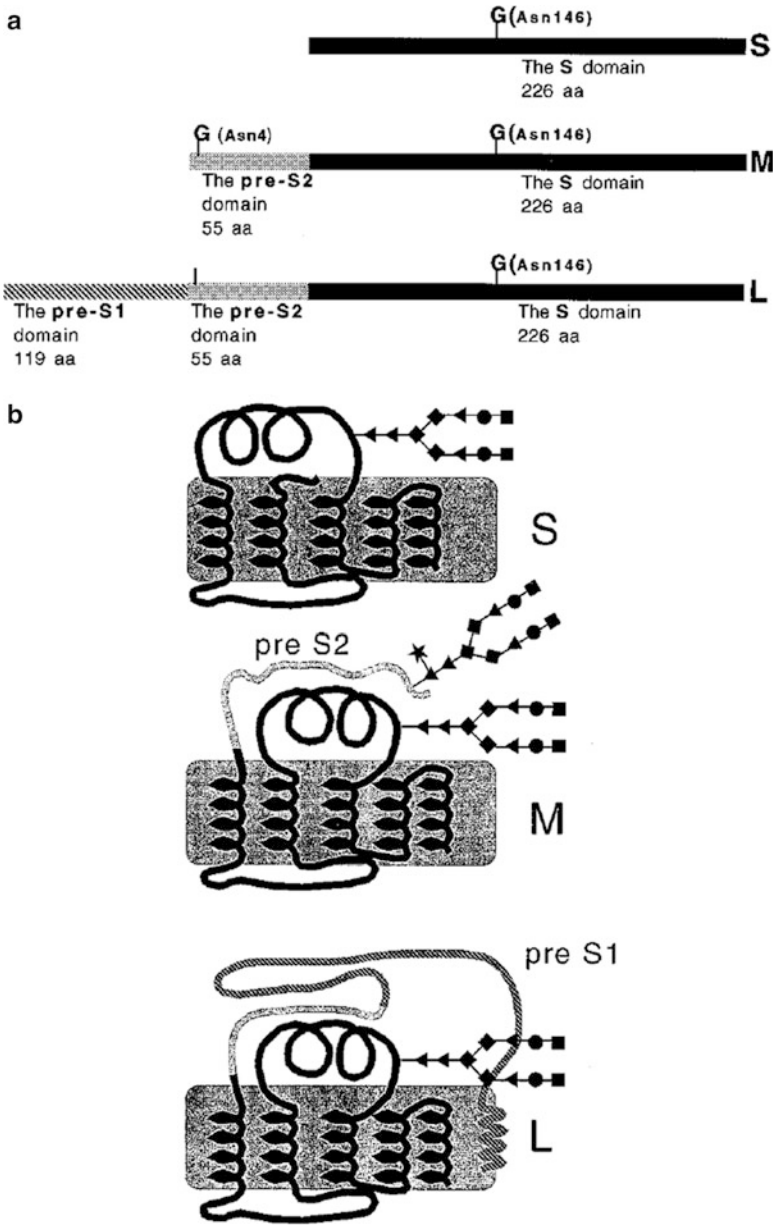


Fig. 13.4 Maturation of hepatitis B virus (HBV) envelope glycoproteins. (**a**) The HBV envelope proteins. All three proteins have a common N-linked glycosylation site at position 146 of the S domain (*marked with a G*). The M protein contains an additional glycan site at amino acid 4 of the pre-S2 domain (*marked with a G*). The L protein, while containing the pre-S2 glycosylation sequon, only utilizes the shared S glycan site (1). (**b**) Structural diagram of the three HBV envelope proteins with attached N-linked glycans. The pre-S1 and pre-S2 domains are indicated. The glycan structures are as follows: mannose (*diamonds*); N-acetylglucosamine (*triangles*); galactose (*circles*); sialic acid (*squares*); fucose (*stars*) (Reproduced from [29]. With permission)

13.6.2 Glycosylation in Orthomyxovirus Replication

Influenza A virus (IAV) is the only species of the Influenza virus A genus belonging to the *Orthomyxoviridae* family that causes annual epidemics and recurring pandemics with potentially severe consequences for public health and global economy [40]. The IAV hemagglutinin (HA) is the virion surface glycoprotein that attaches the virus to glycan receptors on host cell and mediates the fusion of the viral envelope with the membranes of endocytic vesicles to initiate the infectious process (Chap. 16). Moreover, HA is the virion component that stimulates the generation of protective antibodies. All of these important functions of HA are mediated by *N*-linked glycosylation involving attachment of complex glycans to an asparagine residue in a consensus sequence [25].

HA is a homotrimeric integral membrane protein where the monomers are synthesized on membrane-bound ribosomes from a precursor that is then glycosylated and cleaved into two smaller polypeptides: the HA_1 and HA_2 subunits. Each monomer has an ectodomain consisting of a globular head, which harbors the glycan receptor-binding site (HA_1) and a stem region that anchors the protein to the membrane (HA_2) [41]. The IAV HA contains 3–9 N-linked glycosylation sites per subunit, depending on the virus strain (Fig. 13.5). Amino acids sequence analysis has revealed that there is considerable variation in both the number and location of potential glycosylation sites among different HA subtypes and even among variants from a single subtype which is believed to be involved in the evolution of influenza viruses [51]. Variation in protein glycosylation is a more efficient mechanism than the direct mutation of amino acids for the virus to escape the surveillance of the host immune system. This is due to the fact that the glycans are host-derived and hence considered as "self" by the immune system [47]. However, highly conserved glycosylation sites at Asn-12 and Asn-478 (the numbering corresponds to H7) and a further semiconserved site at Asn-28 is [32], the three of them located within the stem region of the HA molecule (Fig. 13.5), have been described to be related to protein folding and maturation [15, 50].

Cleavage of precursor HA by endoproteases occurs late during the transport in the trans-Golgi network. Based on detailed studies employing site-specific mutagenesis on the influenza virus strain A/FPV/Rostock/34 conserved-glycosylation sites of HA and cleavage rates of HA mutants, Roberts and collaborators examined their role in protein transport [36]. It was found that Asn-28 oligosaccharide plays a dominant role in promoting trimerization and proper HA transport since the loss of the carbohydrate at this site interferes with the rapid formation of a transport-competent form of HA early after synthesis. However, only the loss of all three conserved glycosylation sites results in the accumulation of this HA protein at the ER. A highly detailed model for the folding of HA in its natural environment has been proposed indicating that *N*-linked glycans direct the molecular choreography for a ribosome-bound nascent chain as it emerges mediating interactions with chaperones and a foldase, ERp57 [10]. Studies reflecting the impact of each HA glycosylation site on production of the mature infective virus progeny are still

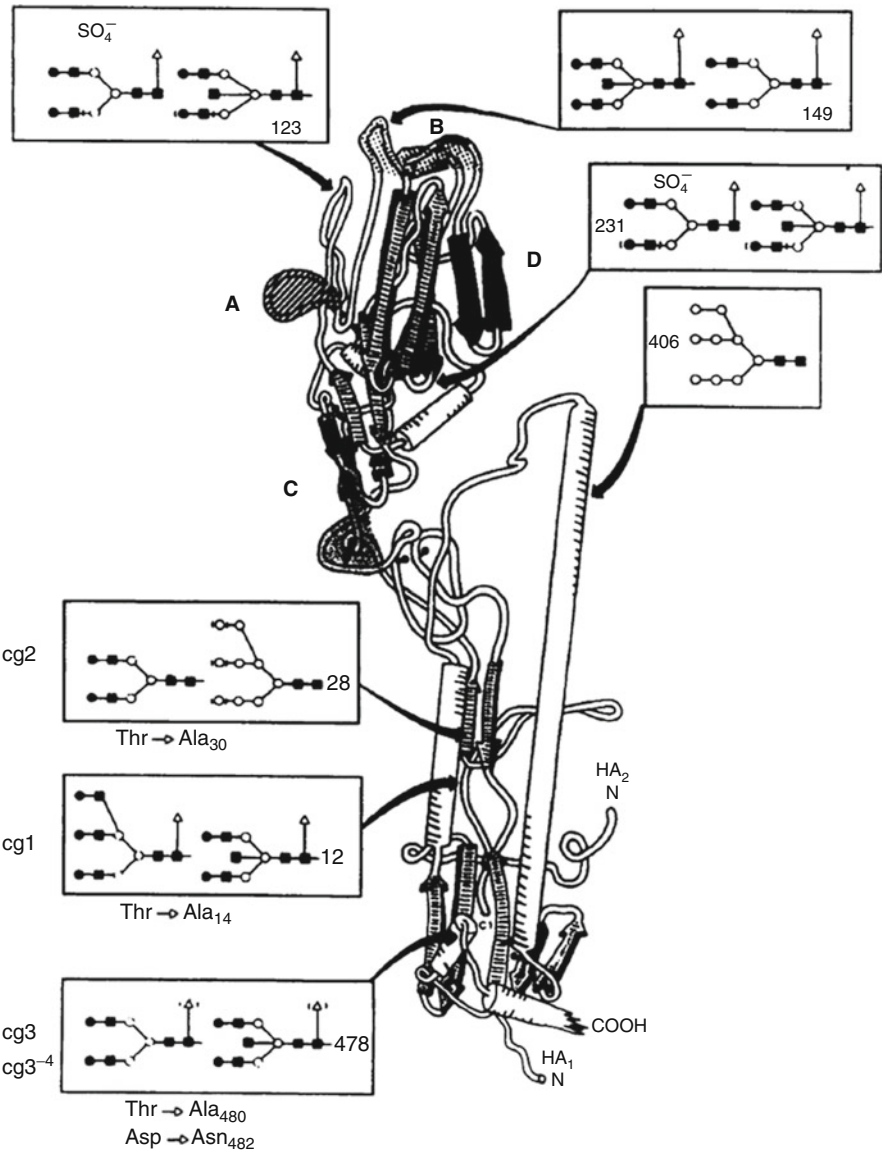


Fig. 13.5 Structural model of the HA monomer; the carbohydrate structures present at the seven glycosylation sites of the HA of influenza virus strain A/FPV/Rostock/34 (H7N1) [50] are shown. The positions of the asparagine residues are indicated in accordance with the H7 HA amino acid sequence. The conserved glycosylation sites are designated cgl (Asn-12) and cg2 (Asn-22) in the HA, subunit and cg3 (Asn-478) in the HA2 subunit and correspond to H3 numbering as residues 22, 38, and 483, respectively (Reproduced from [36]. With permission)

lacking, but it is tempting to hypothesize that correct folding, transport and maturation of HA is a central issue in this regard.

Finally, the amino acid sequence of the HA and, hence, the location of its *N*-linked oligosaccharides are determined by the viral genome, which is replicated by a virus-encoded RNA-dependent RNA polymerase. This enzyme lacks editing functions, thus mutations in all of the viral genes occur at a high frequency. On the other hand, the composition and structure of the oligosaccharides put onto the HA at the various sites is determined by biosynthetic and trimming enzymes provided by the host cell. Thus, the plasticity of the viral genome and the host specificity of the glycosylation machinery can, together, create virus populations that are more heterogeneous in structure and function than those potentially developed by either process alone. This diversity is considered to be responsible for the survival of these viruses in a variety of biological niches and for their ability to overcome the inhibitory effects of neutralizing antibodies and antiviral agents. The glycoprotein nature of the viral HA is therefore a key factor in enabling these viruses to retain their prevalent position amongst the re-emerging human infections.

13.7 Virus Polyhedra: Virus-Derived Assemblages for Long-Term Survival

A small number of viruses belonging to different families undergo a maturation step that is subsequent to the assembly of infectious particles. During the last phase of the replication cycle of these viruses, a fraction of the newly assembled particles are embedded into proteinaceous superstructures, known as polyhedra or occlusion bodies, formed by virus-encoded proteins. Virions trapped inside polyhedra are known as occluded viruses. Occlusion bodies are released from infected cells and play a critical role ensuring the survival of occluded viruses for very long periods (several years) of time after their discharge from the infected organism into the environment. Although the presence of occluded viruses has been mainly associated to three families of insect viruses namely baculoviruses [37], cypoviruses [33], and entomopoxviruses [2], they have also been found in viruses infecting other animal species, *e.g.*, cowpox virus, ectromelia virus, and raccoonpox virus belonging to the poxvirus family that infect different mammal species [22], as well as in different plant viruses [27].

13.7.1 Baculovirus Polyhedra

The *Baculoviridae* is a large family of viruses infecting arthropods mainly belonging to the insect orders *Lepidoptera*, *Diptera* and *Hymenoptera*. Baculovirus virions are rod-shaped enveloped particles enclosing a large double stranded DNA

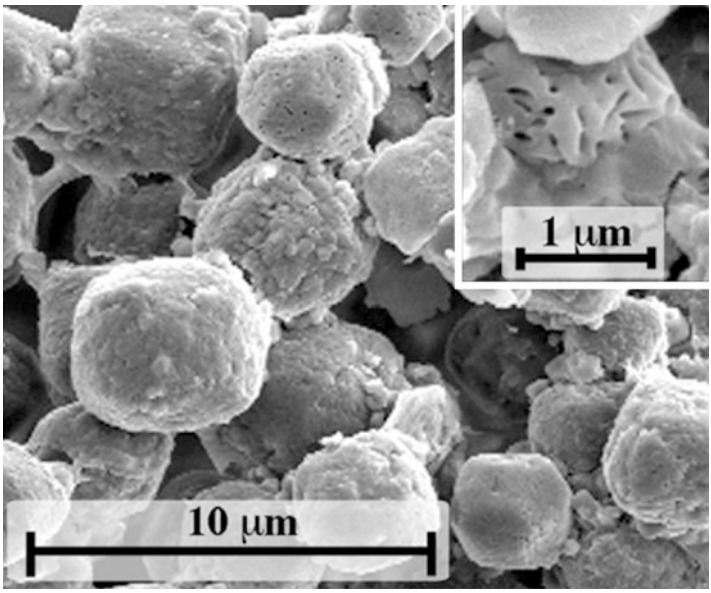


Fig. 13.6 Scanning electron microscopy of polyhedra purified from larvae of porina moths (*Wiseana* spp.) infected with *Wiseana* nucleopolyhedrovirus (WNPV). The *inset* shows an enlarged image area in which imprints of WNPV particles are apparent (Reproduced from [9]. With permission)

(88–160-kbp) genome [20]. Baculovirus polyhedra have been extensively characterized. Two baculovirus genera occlude their virions into protein crystals (Fig. 13.6). The Alphabaculoviruses form a single nuclear polyhedra embedding up to 200 virions surrounded by an electron-dense carbohydrate layer, known as calyx, containing a virus-encoded 34-kDa polypeptide [49]. In contrast, members of the Betabaculovirus genus, generally known as granuloviruses, occlude their virions into smaller cytoplasmic polyhedra usually containing a single virus particle [37]. The matrix of occlusion bodies is composed by a single protein, termed polyhedrin, with a molecular size of 29-kDa. This protein is one of the best-conserved polypeptides amongst the *Baculoviridae* family with a sequence identity of ca. 90 % within Alphabaculovirus, 60 % within Betabaculovirus, and 50 % between the two groups. The atomic structure of the *Autographa californica* nucleopolyhedrovirus (AcMNPV), a Betabaculovirus, has been recently solved [9]. The polyhedrin molecule folds into a central β-sandwich core with an extended N-terminal α-helical projection and a long C-terminal tail. Polyhedrin molecules assemble into trimers where molecules are held together predominantly by hydrophobic interactions. Polyhedra are built by tetrahedral clusters formed by four polyhedrin trimeric units. Tetrahedral polyhedrin clusters are densely packed and stabilized by the docking of C-terminal molecular arms that fit into cavities of neighboring clusters [9].

Polyhedra exhibit an extremely high resistance to severe thermal and chemical conditions. However, they are readily dissolved in alkaline solutions that resemble conditions of the insect intestinal tract. Ingestion of polyhedra by susceptible hosts triggers the release of the occluded viral particles, thus facilitating the infection of midgut cells. Early during infection, non-occluded virions are released from infected midgut cells into the hemolymph promoting the establishment of a systemic infection. The formation of polyhedra takes place during the final stages of the virus replication cycle. Despite the wealth of structural and functional information, mechanisms governing both polyhedra assembly and incorporation of virus particles into growing baculovirus and cypovirus polyhedra remain poorly understood.

13.8 Perspectives and Conclusions

Virus morphogenesis is initiated by the establishment of weak interactions amid their structural subunits. This strategy secures the proper assembly of the virus building blocks but results in the formation of immature assembly products, known as provirions, completely lacking the capacity to initiate a productive infection. The maturation of provirions is an essential and ubiquitous process that transforms fragile, noninfectious assemblages into fully infectious virus particles. This process involves the proteolytic processing and rearrangement of structural elements that consolidate the structure of the virus particle.

The maturation process of enveloped viruses involves a further step consisting on the proteolytic cleavage of glycosylated envelope polypeptides. This event is essential to expose glycoprotein domains involved in the interaction of virus particles with specific cell receptors, and the subsequent fusion of their envelope with the host cell membrane.

Finally, a reduced number of viruses undergo and additional maturation step that results in the incorporation of a fraction of the newly assembled virions into a virus-derived proteinaceous matrix. This superstructures, known as occlusion bodies, play a key role ensuring the survival of infective particles for prolonged periods of time under harsh environmental conditions.

Acknowledgements We thank Dolores Rodríguez and Ana Oña for critical readings of the manuscript. This work was supported by grant from the Spanish Ministry of Economy and Competitiveness AGL2011-24758 to JFR.

References and Further Reading

1. Agrawal DK, Johnson JE (1995) Assembly of the T = 4 Nudaurelia capensis omega virus capsid protein, post-translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system. Virology 207:89–97
2. Bilimoria SL, Arif BM (1979) Subunit protein and alkaline protease of entomopox virus spheroids. Virology 96:596–603

3. Briggs JA, Kräusslich HG (2011) The molecular architecture of HIV. J Mol Biol 410:491–500
4. Brown JC, Newcomb WW (2011) Herpesvirus capsid assembly: insights from structural analysis. Curr Opin Virol 1:142–149
5. Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell 89:263–273
6. Chan WE, Chen SS (2006) Downregulation of human immunodeficiency virus type 1 Gag expression by a gp41 cytoplasmic domain fusion protein. Virology 348:418–429
7. Chertova E, Chertov O, Coren LV, Roser JD, Trubey CM, Bess JW Jr, Sowder RC 2nd, Barsov E, Hood BL, Fisher RJ, Nagashima K, Conrads TP, Veenstra TD, Lifson JD, Ott DE (2006) Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. J Virol 80:9039–9052
8. Chiu E, Coulibaly F, Metcalf P (2012) Insect virus polyhedra, infectious protein crystals that contain virus particles. Curr Opin Struct Biol 22:234–240
9. Coulibaly F, Chiu E, Gutmann S, Rajendran C, Haebel PW, Ikeda K, Mori H, Ward VK, Schulze-Briese C, Metcalf P (2009) The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. Proc Natl Acad Sci U S A 106:22205–22210
10. Daniels R, Kurowski B, Johnson AE, Hebert DN (2003) N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. Mol Cell 1179–1190
11. Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellett PE, Roizman B, Studdert MJ, Thiry E (2009) The order Herpesvirales. Arch Virol 154:171–177
12. Ganem D, Prince AM (2004) Hepatitis B virus infection–natural history and clinical consequences. N Engl J Med 350:1118–11129
13. Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) The structural biology of HIV assembly. Curr Opin Struct Biol 18:203–217
14. Gao M, Matusick-Kumar L, Hurlburt W, DiTusa SF, Newcomb WW, Brown JC, McCann PJ 3rd, Deckman I, Colonno RJ (1994) The protease of herpes simplex virus type 1 is essential for functional capsid formation and viral growth. J Virol 68:3702–3712
15. Gething MJ, McCammon K, Sambrook J (1986) Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. Cell 46:939–950
16. Hanzlik TN, Gordon KH (1997) The tetraviridae. Adv Virus Res 48:101–168
17. Hebert DN, Foellmer B, Helenius A (1995) Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. Cell 81(3):425–433
18. Heermann KH, Goldmann U, Schwartz W, Seyffarth T, Baumgarten H, Gerlich WH (1984) Large surface proteins of hepatitis B virus containing the pre-s sequence. J Virol 52:396–402
19. Helgstrand C, Munshi S, Johnson JE, Liljas L (2004) The refined structure of nudaurelia capensis omega virus reveals control elements for a T = 4 capsid maturation. Virology 318:192–203
20. Herniou EA, Jehle JA (2007) Baculovirus phylogeny and evolution. Curr Drug Targets 8:1043–1050
21. Heymann JB, Cheng N, Newcomb WW, Trus BL, Brown JC, Steven AC (2003) Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. Nat Struct Biol 10:334–341
22. Howard AR, Weisberg AS, Moss B (2010) Congregation of orthopoxvirus virions in cytoplasmic a-type inclusions is mediated by interactions of a bridging protein (A26p) with a matrix protein (ATIp) and a virion membrane-associated protein (A27p). J Virol 84:7592–7602
23. Huovila AP, Eder AM, Fuller SD (1992) Hepatitis B surface antigen assembles in a post-ER, pre-Golgi compartment. J Cell Biol 118:1305–1320
24. Johnson JE (2010) Virus particle maturation: insights into elegantly programmed nanomachines. Curr Opin Struct Biol 20:210–216
25. Klenk HD, Wagner R, Heuer D, Wolff T (2002) Importance of hemagglutinin glycosylation for the biological functions of influenza virus. Virus Res 82:73–75
26. Lu X, Mehta A, Dadmarz M, Dwek R, Blumberg BS, Block TM (1997) Aberrant trafficking of hepatitis B virus glycoproteins in cells in which N-glycan processing is inhibited. Proc Natl Acad Sci U S A 94:2380–2385

27. Martelli GP, Russo M (1977) Plant virus inclusion. Adv Virus Res 21:175–266
28. Matsui T, Lander G, Johnson JE (2009) Characterization of large conformational changes and autoproteolysis in the maturation of a T = 4 virus capsid. J Virol 83(2):1126–1134
29. Mehta A, Lu X, Block TM, Blumberg BS, Dwek RA (1997) Hepatitis B virus (HBV) envelope glycoproteins vary drastically in their sensitivity to glycan processing: evidence that alteration of a single N-linked glycosylation site can regulate HBV secretion. Proc Natl Acad Sci U S A 94:1822–1827
30. Merry T, Astrautsova S (2010) Alternative approaches to antiviral treatments: focusing on glycosylation as a target for antiviral therapy. Biotechnol Appl Biochem 56:103–109
31. Munshi S, Liljas L, Cavarelli J, Bomu W, McKinney B, Reddy V, Johnson JE (1996) The 2.8 A structure of a T = 4 animal virus and its implications for membrane translocation of RNA. J Mol Biol 261:1–10
32. Nobusawa E, Aoyama T, Kato H, Suzuki Y, Tateno Y, Nakajima K (1991) Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza a viruses. Virology 182:475–485
33. Payne CC, Mertens PPC (1983) Cytoplasmic polyhedrosis viruses. In: Joklik WK (ed) The reoviridae. Plenum Press, New York
34. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA, Swanstrom R (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectiousvirions. J Virol 68:8017–8027
35. Rixon FJ (2008) A good catch: packaging the virus genome. Cell Host Microbe 3:120–122
36. Roberts PC, Garten W, Klenk HD (1993) Role of conserved glycosylation sites in maturation and transport of influenza a virus hemagglutinin. J Virol 67:3048–3060
37. Rohrmann GF (1986) Polyhedrin structure. J Gen Virol 67:1499–1513
38. Schmitt S, Glebe D, Alving K, Tolle TK, Linder M, Geyer H, Linder D, Peter-Katalinic J, Gerlich WH, Geyer R (1999) Analysis of the pre-S2 N- and O-linked glycans of the M surface protein from human hepatitis B virus. J Biol Chem 274:11945–11957
39. Schreiber G, Keating AE (2011) Protein binding specificity versus promiscuity. Curr Opin Struct Biol 21:50–61
40. Shinya K, Makino A, Kawaoka Y (2010) Emerging and reemerging influenza virus infections. Vet Pathol 47:53–57
41. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69:531–569
42. Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF (2005) Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr Opin Struct Biol 15:227–236
43. Stiasny K, Fritz R, Pangerl K, Heinz FX (2011) Molecular mechanisms of flavivirus membrane fusion. Amino Acids 41:1159–1163
44. Taylor DJ, Krishna NK, Canady MA, Schneemann A, Johnson JE (2002) Large-scale, pH-dependent, quaternary structure changes in an RNA virus capsid are reversible in the absence of subunit autoproteolysis. J Virol 76:9972–9980
45. Telesnitsky A (2010) Retroviruses: molecular biology, genomics and pathogenesis. Future Virol 5:539–5343
46. Thomsen DR, Roof LL, Homa FL (1994) Assembly of herpes simplex virus (HSV) intermediate capsids in insect cells infected with recombinant baculoviruses expressing HSV capsid proteins. J Virol 68:2442–2457
47. Vigerust DJ, Shepherd VL (2007) Virus glycosylation: role in virulence and immune interactions. Trends Microbiol 15:211–218
48. Werr M, Prange R (1998) Role for calnexin and N-linked glycosylation in the assembly and secretion of hepatitis B virus middle envelope protein particles. J Virol 72:778–782
49. Whitt MA, Manning JS (1988) A phosphorylated 34-kDa protein and a subpopulation of polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus occlusion body. Virology 163:33–42

50. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289:366–373
51. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B (2004) Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology 14:1229–1246

Further Reading

References [8, 13, 21, 24, 35, 41, 42, 43, and 47] *listed above are especially recommended for further reading.*

Chapter 14
Virus Morphogenesis in the Cell: Methods and Observations

Cristina Risco and Isabel Fernández de Castro

Abstract Viruses carry out many of their activities inside cells, where they synthesise proteins that are not incorporated into viral particles. Some of these proteins trigger signals to kidnap cell organelles and factors which will form a new macro-structure, the virus factory, that acts as a physical scaffold for viral replication and assembly. We are only beginning to envisage the extraordinary complexity of these interactions, whose characterisation is a clear experimental challenge for which we now have powerful tools. Conventional study of infection kinetics using virology, biochemistry and cell biology methods can be followed by genome-scale screening and global proteomics. These are important new technologies with which we can identify the cell factors used by viruses at different stages in their life cycle. Light microscopy, electron microscopy and electron tomography, together with labelling methods for molecular mapping *in situ*, show immature viral intermediates, mature virions and recruited cell elements in their natural environment. This chapter describes how these methods are being used to understand the cell biology of viral morphogenesis and suggests what they might achieve in the near future.

Keywords Virus factory • Virus assembly • Virus morphogenesis • Genomics • Proteomics • Light microscopy • Electron microscopy • Tomography • 3D electron microscopy • Super-resolution microscopy • Correlative microscopy • Molecular mapping • Immunofluorescence • Immuno gold • Immuno electron microscopy • Clonable tags

C. Risco (✉) • I. Fernández de Castro
Cell Structure Laboratory, Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: crisco@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_14,

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ASFV	African swine fever virus
CLEM	Correlative light and electron microscopy
DNA	Deoxyribonucleic acid
ESCRT	Endosomal sorting complex required for transport
ET	Electron tomography
GFP	Green fluorescent protein
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
IEM	Immunoelectron microscopy
IF	Immunofluorescence
LM	Light microscopy
RC	Replication complex
RNA	Ribonucleic acid
SEM	Scanning electron microscopy
TEM	Transmission electron microscopy
VV	Vaccinia virus
Y2H	Yeast two-hybrid

14.1 Introduction: Cell Biology of Virus Morphogenesis and the Concept of the Virus Factory

The idea of viruses as inert molecular entities has progressively been transformed since scientists began to discover the myriad of interactions that occur during the intracellular phase of virus life. Restricted by their own limited genetic repertoire, viruses need to use a number of cell factors for genome replication and morphogenesis. Identification of these factors is essential for understanding the virus morphogenetic processes that often take place in intracellular structures known as viral inclusions, virosomes or viral factories. Viruses are generally thought to build these structures to recruit and concentrate viral and cell factors needed for replication and assembly; they are able to modify a variety of cell organelles and to create new inter-organelle contacts (Fig. 14.1). In the mammalian cell, virus assembly can start inside the nucleus, in association with components of the secretory pathway (Golgi, endoplasmic reticulum or ER), at different points in the endocytic pathway (endosomes, multivesicular bodies or MVB) or at the plasma membrane. Mitochondria and cytoskeletal elements are present in the factories built by many different viruses.

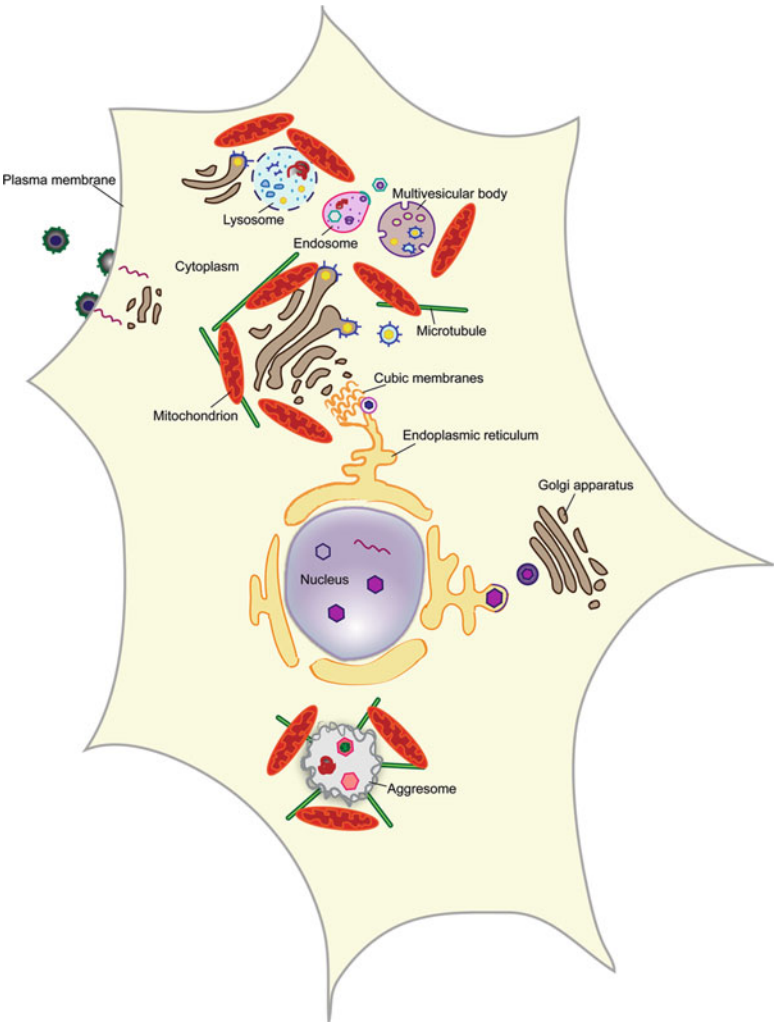


Fig. 14.1 Cell organelles used for viral morphogenesis. The cell nucleus is used by herpes- and papillomaviruses; corona-, bunya- and flaviviruses use components of the secretory pathway such as the endoplasmic reticulum and Golgi; togaviruses and the human cytomegalovirus use components of the endo-lysosomal pathway such as endosomes, multivesicular bodies and lysosomes; the African swine fever virus (ASFV) and poxviruses assemble in aggresome-like structures; assembly sites of retroviruses at the plasma membrane might be connected with a cytoplasmic factory. Viruses modify endomembranes, recruit mitochondria and cytoskeleton, and create new inter-organelle contacts

Virus factories are usually rather large, several microns in diameter, and are very dynamic, as their structure changes over time in accordance with virus needs [1]. This strategy appears to allow viruses to generate progeny with maximum efficiency in the use of cell resources. It is also thought that viruses can use these

scaffolds to protect themselves from cellular antiviral defences [2]. Viral factories were first described for large DNA viruses such as the African swine fever virus (ASFV) and the poxviruses; we now know that many DNA and RNA viruses build factories in the cytosol, in the nucleus, or both. Due to our limited knowledge of cell nucleus organisation, characterisation of nuclear viral factories is only beginning. In contrast, viral factories assembled in the cytosol are being studied extensively and there are many reports on their architecture and activities. Inside these cytosolic factories, morphogenesis of new viruses begins with the transport of replicated genomes from the structures that harbour the replication complexes (RC) to the assembly sites. Depending on the virus, a number of maturation steps will follow (see Chap. 13 for details) before new viral particles are ready for delivery and propagation.

This chapter will describe fundamental tools for diving inside the infected cell and understanding virus assembly. We also describe relevant examples of virus-cell interactions during virus morphogenesis that are being discovered using *in situ* techniques.

14.2 Methods for the Study of Virus-Cell Interactions During Morphogenesis

By identifying essential cell factors, studies based on recent advances in molecular biology, genomics and proteomics are broadening our knowledge of viral morphogenesis. To confirm and understand the role of potentially interesting genes in viral assembly, the information acquired must be subsequently analysed at a more complex level in infected cells. Exploring existing databases on cell pathways which combine information on genetic, metabolic and signal networks based on the literature can be a first step in further work that on many occasions will also include imaging with a variety of microscopy technologies.

14.2.1 Identification of Essential Cell Factors

Characterisation of infection kinetics is the first step in the study of viral morphogenesis. Conventional virology, biochemistry and cell biology methods allow us to determine optimal experimental conditions, including times post-infection (p.i.) and the most appropriate cell types. Key factors can be identified using two main groups of methods, (1) those that analyze gene expression patterns, and (2) those that study protein-protein interactions.

A number of new techniques can be applied to study interactions between viruses and cells on a genomic scale [3]. Microarrays are being used to identify mRNA transcription patterns in different phases of the virus life cycle. DNA gene

chips detect differences in gene expression between uninfected and virus-infected cells and at different infection stages. Gene data bases are then useful for associating the genes identified with specific cell pathways [4, 5]. High-throughput screening based on RNA interference (RNAi) is another category of methods that analyse gene expression patterns. RNAi is an RNA-dependent gene silencing process within living cells that is often exploited to study the function of genes. This emerging technology is used to study how viruses interact with their hosts at the molecular level. Analysis at various times post-infection has identified a number of cell factors potentially involved in viral morphogenesis, for example for dengue [6], influenza [7] and retroviruses [8]. RNAi may be used for large-scale screens that systematically shut down each gene in the cell. Studies using this approach have shown a requirement for cell factors such as the ESCRT machinery for assembly of the human immunodeficiency virus, HIV [9].

The group of techniques termed proteomics includes powerful methods to study protein-protein interactions. The yeast two-hybrid (Y2H) assay system remains one of the most amenable techniques and is widely used to search for virus-host interactions. Y2H works by expressing two candidate proteins in the yeast cell. Bait and prey proteins are fused either to a promoter-specific DNA-binding domain or to a transcription activation domain. Interaction between the two proteins in the yeast nucleus brings both domains together so that they can initiate expression of a reporter gene [10]. Individual bait proteins can be screened for interaction with a library of prey proteins. Genome-scale Y2H studies were used to identify 314 virus-host interactions for HCV [11] 109 interactions for vaccinia virus [12] and nine for HIV-1 [13]. Similar techniques, such as the yeast three-hybrid system, can be used to study interactions between nucleic acids and proteins [14, 15]. This group of methods also includes pull-down and immunoprecipitation assays, and tandem affinity purification (TAP) tagging approaches, as well as protein identification by quantitative and semi-quantitative mass spectrometry. Whereas Y2H usually detects transient interactions, affinity-tag purification mass spectrometry shows stable, stoichiometric complexes. Since cell proteins often incorporate into viral particles, these techniques can be applied to the study of protein-protein interactions in the infected cell and in purified viral assembly intermediates [16]. This is the case of clathrin, for example, which was found in retrovirus particles; clathrin was only recently identified as one of the cell factors that facilitate accurate morphogenesis of several retroviruses [17]. Y2H technology also detected cell proteins that interfere with virus assembly and viral proteins that block them. This is the case of tetherin, first detected by proteomics and mass spectrometric protein identification as a cell factor that restricts retrovirus assembly [18], and later confirmed as a restriction factor for a wide variety of enveloped viruses [19]. Viruses have several anti-tetherin proteins to counteract the effect of this factor [20]. The current challenge of high-throughput technologies is to develop more efficient informatics tools to accurately analyse the vast amount of information they can provide [21].

The next step in the characterisation of in-cell virus assembly pathways would be to study the specific roles of the factors identified in their natural environment; to

do this, we must of course return to the infected cell. Functions can be tested by protein depletion or overexpression, by mutagenesis or by protein targeting with tags. To visualise key factors in infected cells, there are a variety of classical and novel microscopy techniques that will be described in the following sections.

14.2.2 Studying Viral Morphogenesis in Situ with Light and Electron Microscopy

Microscopy has played an essential role in our understanding of cell architecture and viral assembly (see Chap. 3). Light microscopy (LM) and transmission electron microscopy (TEM) provide different types of information about viral infection, ranging from general events that involve the whole cell to the detailed imaging of nascent and maturing viral particles in specific cell compartments [22] (Fig. 14.2). With resolutions in the 100–500 nm range, LM shows organelle recruitment and modification in the assembly compartment and, in the case of the largest viruses, individual new viral particles as well [23]. In immunofluorescence assays using antibodies to viral proteins and cell compartments, we can see where viral structural and scaffolding proteins (Chap. 11) concentrate to create the assembly sites. Light microscopy shows, for example, that the same virus can build distinct factories in different cell types, depending on specific characteristics of the cell (Fig. 14.2a, b). Functional viruses that express proteins fused with clonable fluorescent tags such as the green fluorescent protein (GFP) (see Sect. 14.3 for details) can be followed in live cells [24]. Video microscopy facilitates dynamic characterisation of the biogenesis of the factory and virus assembly in real time. Time course experiments with antibody-labelled permeabilised cells or video microscopy studies with GFP fusions in living cells are essential for selecting specific conditions, such as the best times p.i., for more detailed, higher resolution study by TEM.

Electron microscopy generally uses ultra-thin sections of cells previously embedded in plastic resins after conventional fixation and dehydration; alternatively, cells are processed at low temperature after or upon fixation for optimal preservation of ultrastructure (see Chap. 3 for sample preparation details), or subjected to criofixation procedures such as high-pressure freezing prior to freeze substitution and embedding, or freeze-fracture analyses. With resolutions in the range of a few nanometers, cell TEM can show changes in shape and size of virus assembly intermediates in specific intracellular compartments [25, 26] (Fig. 14.2c, d). TEM of infected cells shows that mitochondria, endomembranes and cytoskeleton often participate in organising the structure that supports viral assembly. To complete maturation and become infectious, immature viral particles must often travel within the factory to find specific cell factors. At late times post-infection, once virus progeny have been produced and must find their way out of the cell, the factory can be dismantled [1].

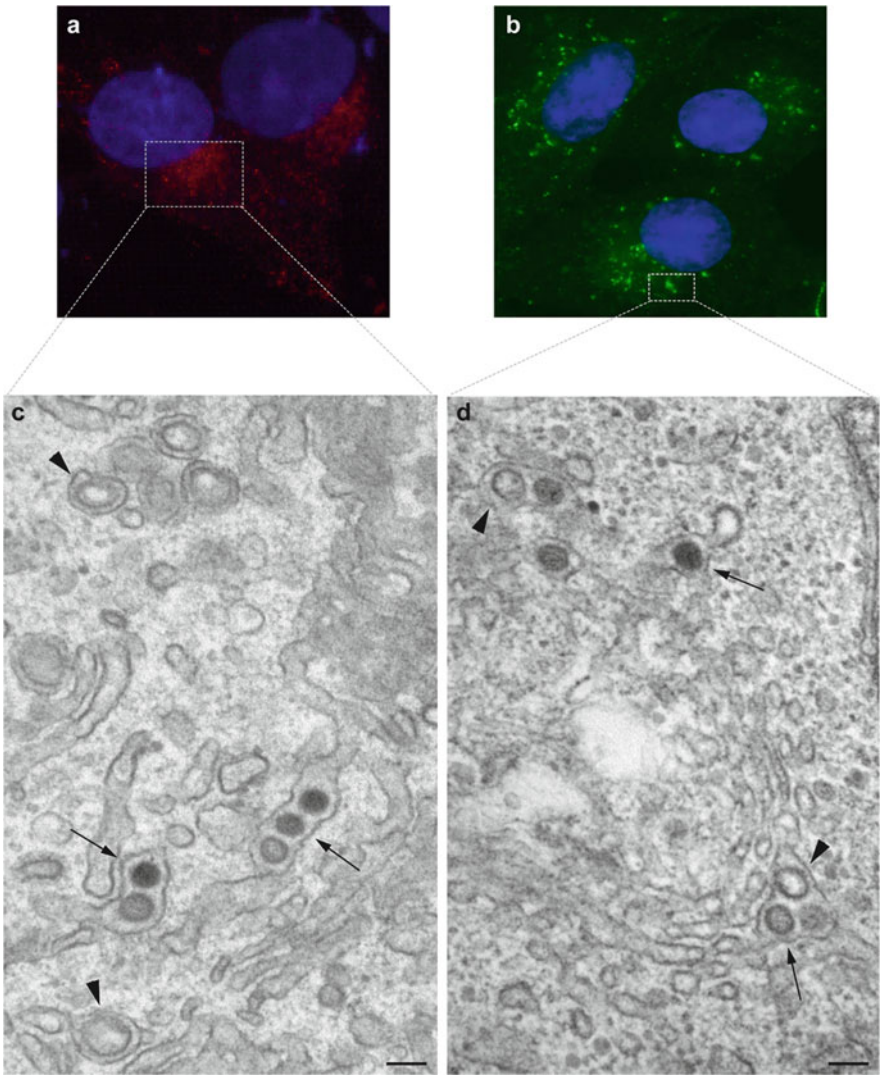


Fig. 14.2 Studying virus assembly by light and electron microscopy. (**a**, **b**) Immunofluorescence microscopy of BHK-21 (**a**) and Vero cells (**b**) infected with a bunyavirus at 10 h p.i. (h.p.i.). Cells were labelled with an antibody specific for one of the viral structural proteins that concentrates at the assembly sites. A single large perinuclear factory is formed in BHK-21 cells, whereas many mini-factories are seen in Vero cells. (**c**, **d**) TEM of BHK-21 (**c**) and Vero cells (**d**) at 10 h.p.i. In both cases, similar spherules, the structures that harbour the RC (*arrowheads*) [22] and viral intermediates (*arrows*) are distinguished in Golgi membranes. Scale bars, 100 nm

TEM of cells in two dimensions provides a first glimpse of the viral morphogenetic pathways *in situ*. The size of cells (several microns in diameter) and the thickness of ultra-thin sections (50–100 nm) nonetheless constitute a major

limitation, because we are actually studying single planes of much larger structures. In these conditions, many elements and contacts can be missed. Three-dimensional TEM avoids this limitation; the variety of methods available is described in the following sections.

14.2.3 Visualising Virus Assembly in Three Dimensions I: 3D Reconstructions of Serial Sections, TEM and SEM

In conventional TEM, electrons must traverse the sample to generate a projection image. Samples must therefore be thin, around 50–100 nm, thinner than many viruses. Given that eukaryotic cells are several microns in diameter, the ultra-thin sections are single planes of much larger structures. For conventional TEM, cells are fixed on culture plates and collected by low speed centrifugation, followed by sectioning of the pellet. In these conditions, cells preserve their morphology but present a variety of orientations (Fig. 14.3a); analysis is therefore restricted to random, unique sections of cells, and scarce or non-randomly distributed elements can be missed completely. Oriented serial sections solve this problem (Fig. 14.3b). By collecting all serial sections from each cell, all intracellular elements can be detected and studied. This strategy guarantees a complete analysis without missing any intracellular event of interest; immunofluorescence can assist by localising where viruses are assembled within a cell and thus, where cells should be sectioned for TEM [22]. The study of oriented serial sections is very informative and has revealed unreported contacts between cell organelles, RC and assembly sites, as well as the relationships of different viral intermediates with specific cell elements (Fig. 14.3c). After image processing and segmentation to assign identities to all individual structures in the images, serial sections can be aligned and combined in 3D reconstructions [27]. With resolution values of ~5 nm in the X and Y axes, these 3D reconstructions can show viruses in factories with considerable detail (Fig. 14.3d, e). Resolution in the Z axis is limited, however, with values of ~50–100 nm due to imperfections in the alignment process.

Scanning electron microscopy (SEM) can also be used to visualise virus morphogenesis. These microscopes use a focused beam of high-energy electrons to generate a variety of signals at the surface of solid specimens. Samples are covered by a thin layer of metal and scanned with a primary source of electrons; secondary electrons are then released from the sample surface and collected by a detector. The signals derived from electron-sample interactions furnish information about the sample, including external morphology in 3D. Modern scanning electron microscopes can now provide resolutions as precise as 2–5 nm, near that achieved by cellular electron microscopy [28]. SEM images can show morphological changes in large virus particles during maturation *in situ*; this is the case, for example, of the giant mimivirus (Fig. 14.4a, b). In lysed cells and isolated virus factories visualised by SEM, immature viruses are clearly distinguished at short

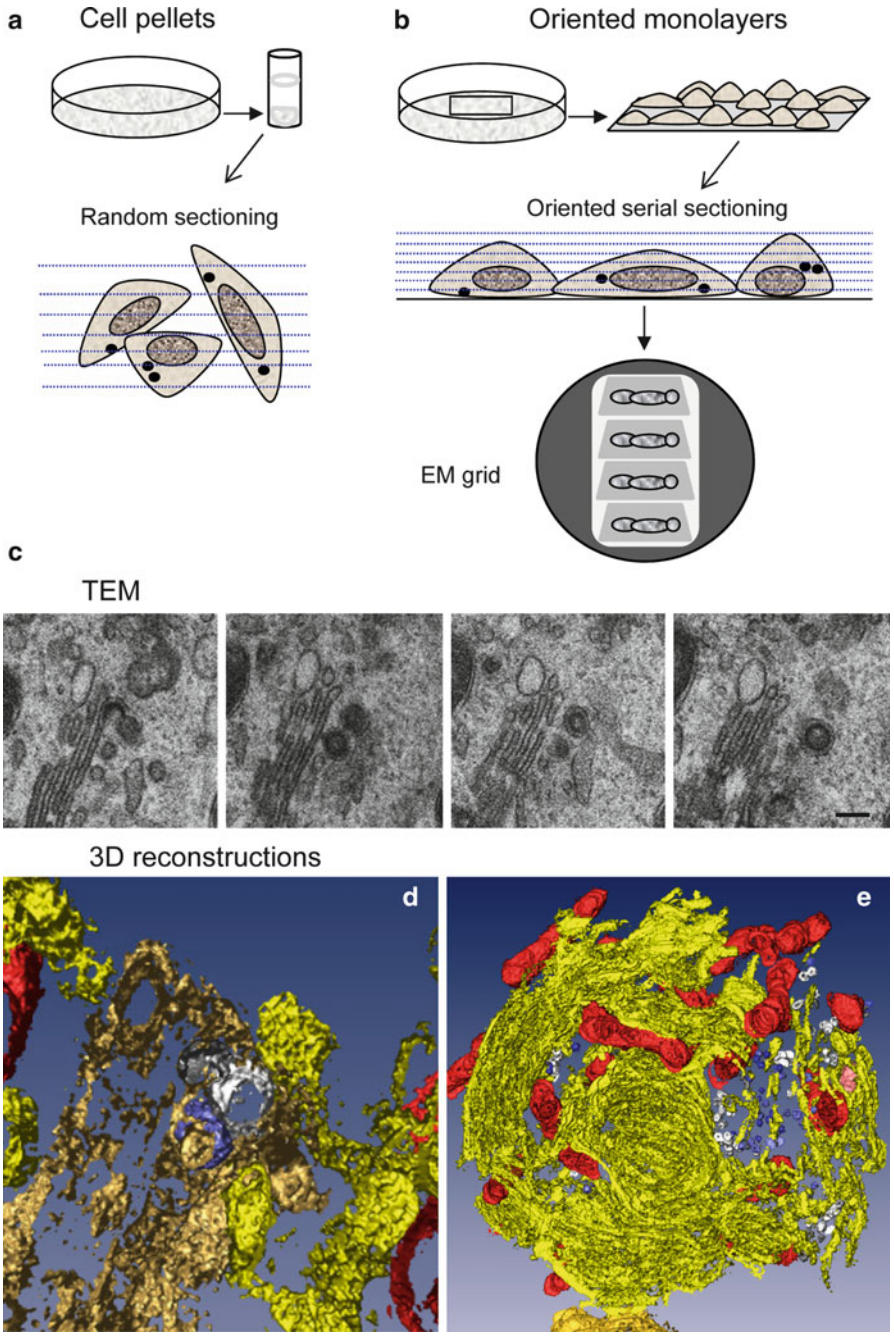


Fig. 14.3 Serial sections, TEM and 3D reconstructions. (**a**, **b**) Summary of the principles and differences between conventional ultra-thin sectioning and oriented serial sectioning. (**c**) Serial sections in TEM and (**d**) 3D reconstruction showing the interaction between a spherule, the structure that harbours the RC (*white*) and a viral particle (*blue*) in Golgi membranes (*beige*). Mitochondria are segmented in *red* and rough endoplasmic reticulum (RER) in *yellow*. (**e**) 3D reconstruction of a viral factory from a different cell. In this case, 15 serial sections were used. The Golgi complex has been removed to improve visualisation of RC and viral particles. Scale bar, 100 nm

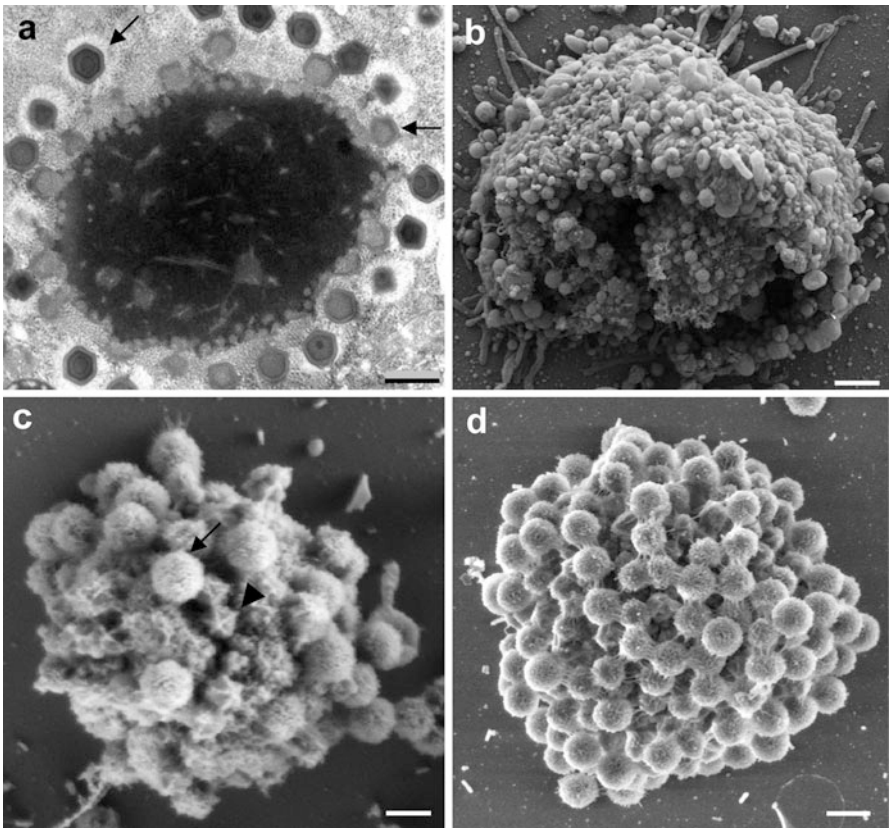


Fig. 14.4 Mimivirus factory in 2D and 3D by TEM and SEM. (**a**) 2D views of a virus factory as visualised by TEM, showing virus particles (*arrows*) at various assembly stages. (**b**) 3D views of a viral factory within an amoeba cell lysed at 8 h.p.i. and visualised by SEM. (**c**) SEM of a factory isolated at 8 h.p.i. Viral particles are seen at various assembly stages. The *arrow* indicates a mature virus particle and the *arrowhead*, an immature particle. (**d**) SEM of a virus factory isolated at 10 h. p.i. Only mature viruses can be detected. Scale bars, 500 nm in (**a**) and (**d**); 2 μm in (**b**); 300 nm in (**c**) (Reproduced with permission from [29])

times p.i. (Fig. 14.4c), whereas mature virions accumulate in the factory at longer times p.i. (Fig. 14.4d) [29].

TEM in combination with freeze-fracture and metal shadowing allows higher resolution 3D views of viruses inside intact cells. Frozen cells are fractured mechanically and covered with a thin layer of platinum and carbon in a vacuum chamber, the metal replicas are then washed to eliminate organic materials and mounted on EM grids. The main limitation of this approach derives from the unpredictable patterns of the fracture planes and the difficulty in interpreting the images. It can nevertheless be useful when studying the morphogenesis of large enveloped viruses (Chap. 11) and their interactions with cell endomembranes during assembly and maturation [30].

14.2.4 Visualizing Virus Assembly in Three Dimensions II: Electron Tomography

Electron tomography (ET) of infected cells shows virus assembly in 3D with resolutions of ~3-5 nm in all three axes, X, Y, and Z (see Chap. 3). Higher resolutions of ~1 nm have been reported, as was the case of the budding HIV viral particles visualised by cryo-electron microscopy on the surface of intact human cells [31]. Both normal and aberrant budding events were visualised on the cell surface, suggesting that cellular and/or viral factors control the quality of virus assembly and maturation. A limitation of cellular tomography is that the maximum thickness of the sample must be <0.5 μm, whereas virus factories are larger; however, conventional TEM methods like those described above can help to select specific elements within cells for more detailed ET structural analysis. Thanks to ET, elusive structures have been visualised for the first time, such as the transfer of viral genomes from RC to assembly sites in dengue virus-infected cells [32] (Fig. 14.5a, b). The morphogenesis of poxviruses is another good example; vaccinia virus (VV) assembly and architecture were the subject of numerous studies, but the organisation and biogenesis of immature and mature VV particles were not understood until the first ET studies were carried out [33, 34]. These analyses showed unprecedented remodelling of cell endomembranes during VV particle assembly (Fig. 14.5c, d). Tomograms are analysed in detail in computational slices of 1–2 nm extracted from the original volumes (Fig. 14.5a). As with serial sections in conventional TEM, the relevant information is contained in these single planes, whereas 3D representations are models used to summarise the most relevant features in the tomograms (see Chap. 3 for details).

In summary, two main groups of methods are available for 3D studies of viral assembly. 3D reconstructions of serial sections by TEM or analysis of surface morphology by SEM will show the general organisation of the virus factory, the inter-organelle contacts, and changes in the cell compartments where viruses are formed. In contrast, electron tomography is more appropriate for studying individual viral particles and to obtain fine details of their maturation *in situ*.

14.3 Molecular Mapping of Viral Morphogenesis

The discipline known as histochemistry includes a great variety of techniques to visualise molecules in biological samples [35]. There are several histochemical methods to localise nucleic acids, lipids, sugars, and other molecules, some of which have been used to label viruses in cells. Nonetheless, protein-labelling techniques are by far the most developed; because of their importance in the study of virus morphogenesis, in this section we will focus on methods that detect proteins in light and electron microscopy.

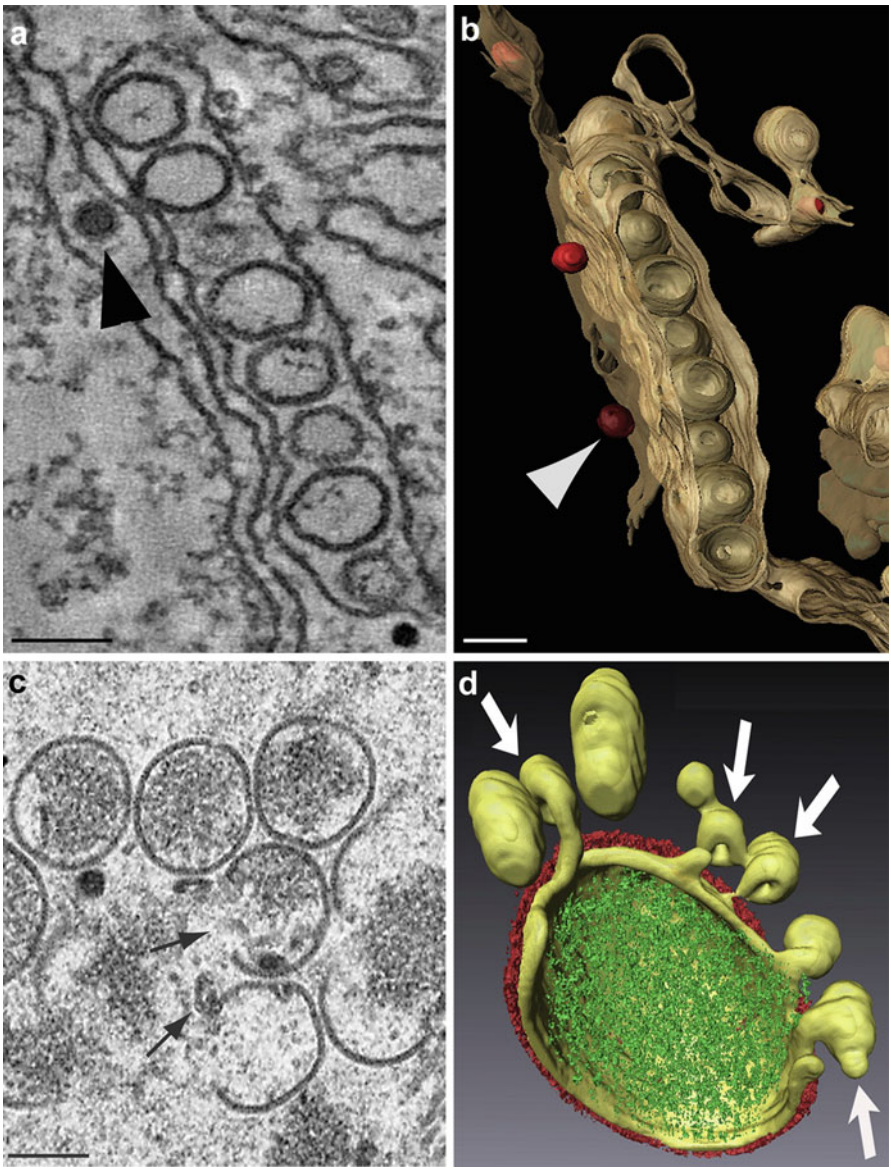


Fig. 14.5 3D electron tomography of virus assembly. (**a**) Computational slice and (**b**) 3D reconstruction of the dengue virus factory as visualised by ET. Nascent viral particles (*arrowheads*) face the spherules that harbour the RC. (**c**, **d**) 2D TEM and 3D ET, respectively, of immature VV particles in the process of assembly from cell membranes (*arrows*). ET shows how the viral envelope is connected to a collection of open membrane structures and how these membranes contribute to envelope formation. Scale bars, 100 nm in (**a**) and (**b**); 250 nm in (**c**) (Reproduced with permission from [32] (**a**) and (**b**), and [33] (**d**))

14.3.1 Antibodies, Immunofluorescence and Immunoelectron Microscopy

Previous sections of this chapter have shown how to identify key factors for virus morphogenesis and how to visualise virus assembly in cells in two and three dimensions. To link the information from both methods and to fully understand viral morphogenesis, we also need methods for molecular mapping *in situ*. Molecules of interest, which are viral and cellular factors, can be localised in cells with specific antibodies in immunolabelling assays or with clonable tags (Fig. 14.6).

Antibodies detect proteins with high specificity and variable sensitivity. Primary antibodies generated against proteins of interest are detected with secondary antibodies conjugated to a fluorescent probe for LM visualisation, or to an electron-dense colloidal gold particle, which is easy to detect by EM in immunogold assays (Fig. 14.6a). Antibodies have been and are still fundamental tools in electron microscopy [36, 37]. When immunolabelling proteins on cryosections, where cells have not been dehydrated and maintain their proteins in a natural hydrated state, the sensitivity of antibody detection is usually higher than that obtained when labelling sections of dehydrated, resin-embedded cells. Since the introduction of cryosectioning, the method has been improved and perfected [38, 39]. This approach allows colocalisation of nascent and maturing viral particles in specific cell compartments and the proteins being incorporated into assembling virus particles (Fig. 14.6b). Although antibodies are usually very specific, information derived from these experiments is later confirmed in biochemical assays.

Due to their large size, antibodies that recognise internal structures must be used on cell sections. Alternatively, cells can be permeabilised to label intracellular compartments (Fig. 14.6c), although this is incompatible with preservation of fine ultrastructure. Certain permeabilisation protocols use the bacterial exotoxin streptolysin O (SLO) to open pores in the plasma membrane while leaving intracellular membranes untouched. This was used in pre-embedding immunogold assays to follow viral and cell proteins during VV assembly from intracellular membranes [40].

An important limitation of antibodies is their variable sensitivity due to epitope loss during sample preparation and to macromolecular interactions inside cells that often mask the protein epitopes *in vivo*. This is particularly problematic when the proteins of interest are part of densely packed macromolecular complexes, such as those involved in virus assembly. The use of clonable tags can overcome these limitations.

14.3.2 Clonable Tags

Jellyfish green fluorescent protein (GFP), its mutants and homologues have caused a true revolution in cell biology. If proteins fused to a fluorescent tag maintain their

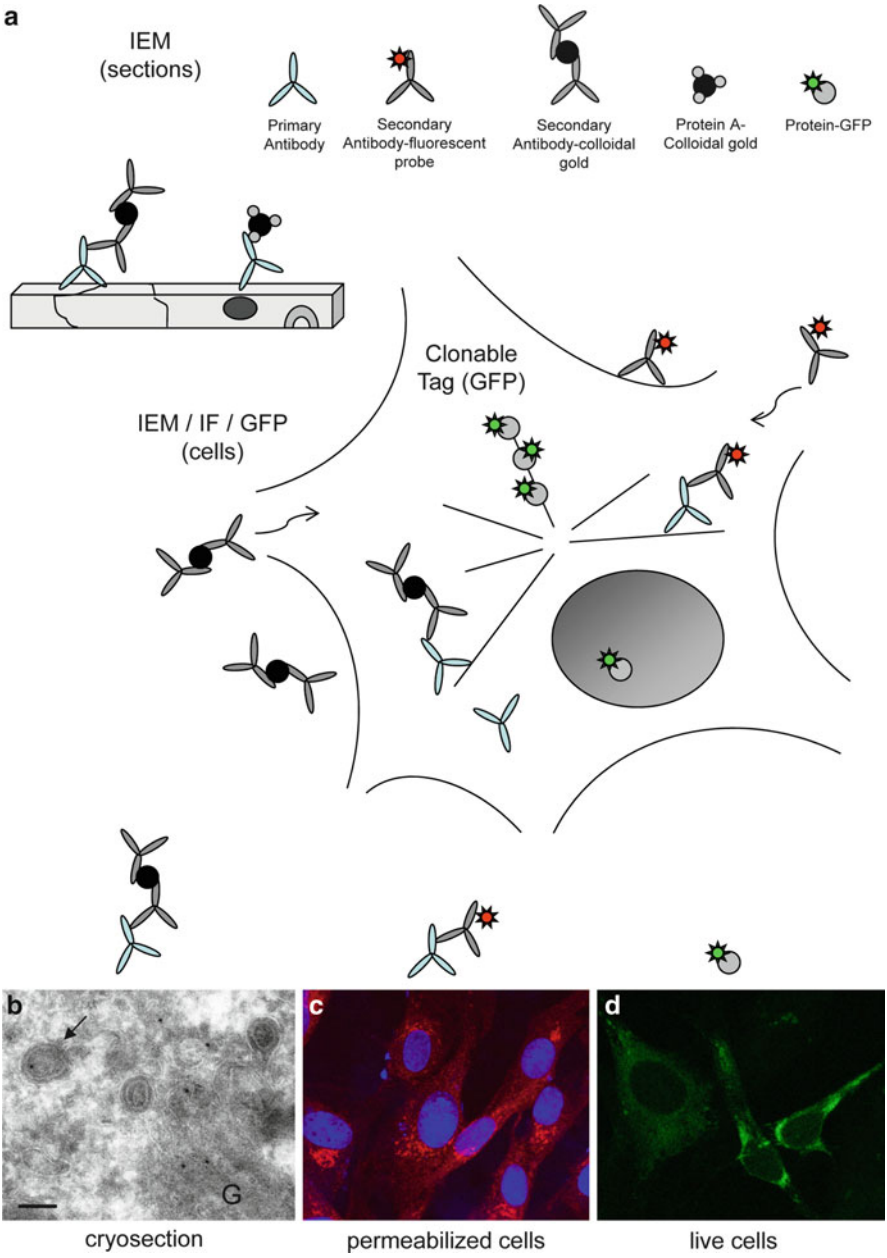


Fig. 14.6 Molecular mapping of virus assembly with antibodies and clonable tags. (**a**) Schemes showing the principles of labelling on thin sections, whole permeabilised cells and intact cells. (**b**) Immunogold detection on a cryosection from a bunyavirus-infected cell. Cells were labelled with a primary antibody specific for a viral scaffolding protein, followed by secondary antibodies conjugated to 10 nm colloidal gold particles. The protein is detected in Golgi membranes (G) and

normal functions, their movements and interactions with intracellular structures can be tracked in living cells (Fig. 14.6d). In conjunction with the new generation of LM methods, we can even follow the assembly of individual virus particles at real time [41]. GFP is a relatively large molecule (27 kDa) and, since viral proteins have very strict structural requirements, fusion to GFP can produce loss of function. If this is the case, probes smaller than GFP could be a viable option. The group of smaller probes includes the fluorescent flavoprotein known as miniSOG, which is less than half the size of GFP [42], or the smaller tetra-cysteine motifs, which are visualised after incubation with fluorescent biarsenical molecules such as ReAsH or FlAsH [43].

The use of clonable tags in TEM and ET would be a considerable advance for high resolution studies of virus assembly. The first clonable tags for EM validated in cells have yet to be used in virus morphogenesis studies, but show good prospects and will be listed in Sect. 14.5.3.

14.4 The Search for Signalling Pathways

The initiation of virus morphogenesis usually requires major reorganisation of cell membrane compartments and the cytoskeleton. Many laboratories are applying the methods described in previous sections to understand how viruses take control of cells and recruit all the necessary materials for their morphogenesis. The complexity of the interactions inside infected cells can be extraordinary, as illustrated by the viruses known as virophages that parasitise the factory built by other viruses. Some images of cells co-infected by the giant mimivirus and the virophage "Sputnik" suggest that the small virophage not only uses the materials recruited by the mimivirus for its own replication, but might even hide inside the mimivirus particles to exit the cell [44]. We still do not know how organelles and materials are recruited or how replicated viral genomes are transported from RC to assembly sites. We do know, however, that viruses target their proteins to specific cell compartments, that some viruses use a cellular defence mechanism termed the aggresome response, and that signals related to organelle movement on cytoskeletal tracks could also be involved [1, 2].

Fig. 14.6 (continued) viral particles (*arrow*). (**c**) Immunofluorescence detection of the same scaffolding protein in permeabilised cells. (**d**) Still image from a video recorded in a fluorescence microscope equipped for live cell imaging. Cells were infected with a recombinant virus that expresses the same scaffolding protein fused with GFP. Scale bar, 100 nm

14.4.1 DNA Viruses and Cell Aggresomes

Aggresomes are a defence response of cells to protein misfolding and aggregation. These inclusions form at the microtubule-organising centre (MTOC), where they enclose potentially toxic protein aggregates within vimentin cages. The large DNA viruses of animals such as the African swine fever virus (ASFV), the poxviruses, and the iridovirus frog virus three build factories that closely resemble cell aggresomes (Fig. 14.7a) [1, 2, 45, 46]. Virus factories and aggresomes both assemble at the MTOC, recruit mitochondria and cell chaperones, build a cage with vimentin filaments, and are maintained by the activity of dynein motors on microtubules. EM shows membranes, ribosomes, viral intermediates and fully assembled viruses inside the vimentin cage of the ASFV aggresome-like factory (Fig. 14.7b). It is suggested that viruses kidnap the aggresome pathway to avoid being recognised as foreign, or alternatively, to be mistaken for a misfolded protein by the cell, thus triggering the aggresome response [2].

A similar strategy might be used by viruses that replicate and assemble in the cell nucleus and associate with structures known as POD (potential oncogenic domains). POD are nuclear aggresomes used by herpes-, papilloma-, adeno- and parvoviruses. Recent observations suggest that some RNA viruses use aggresomes to build their factories. Since the common feature of factories built by RNA viruses is the remodelling and recruitment of cell endomembranes, however, different signalling pathways must be triggered in this case.

14.4.2 RNA Viruses and Membrane Remodelling

DNA viruses usually build distinct, and even distant factories for genome replication and morphogenesis. Herpesviruses, for example, must connect the first steps of assembly inside the nucleus with subsequent incorporation of proteins and membranes in the cytosol; poxviruses must coordinate replication in cytosolic mini-nuclei with primary assembly in aggresome-like structures and final wrapping in the Golgi apparatus [1]. In contrast, RNA viruses often induce the construction of a single sophisticated membranous web in which replication complexes and assembly sites are located near each other. Expression of viral replicase complexes is usually sufficient to trigger membrane remodelling and organelle recruitment [47, 48].

Viruses take control of cell endomembranes by interfering with lipid metabolism, protein regulation and transport. The secretory pathway is the most common target for this virus-induced membrane remodelling, while the endocytic pathway also participates in some cases (Fig. 14.1). The two pathways are closely related, in fact, and converge at the trans-Golgi compartment. Described in the literature with many different names, the membranous tubuloreticular structures (TBS) often detected in infected cells are indeed cubic membranes that consist of highly curved, 3D-folded lipid bilayers. Alterations in cholesterol metabolism are linked to the

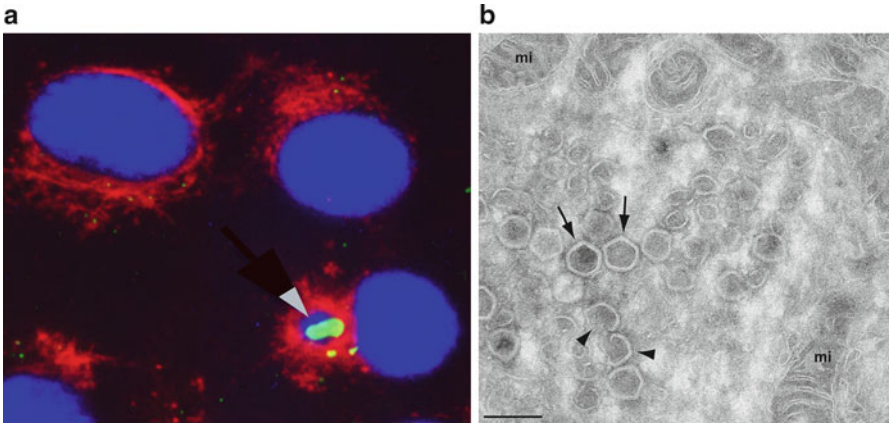


Fig. 14.7 Virus assembly in cell aggresomes. (**a**) The factories of the ASFV resemble cell aggresomes, as shown by immunofluorescence. Viral proteins (*green*) co-localise with DNA (*blue*) inside vimentin cages (*red*). (**b**) TEM of an ASFV factory, showing mitochondria (mi), viral intermediates (*arrowheads*) and mature viruses (*arrows*). Scale bar, 250 nm. Images kindly provided by Dr. Germán Andrés (CBMSO-CSIC, Spain)

biogenesis of cubic membranes, used for assembly by some viruses such as the SARS (severe acute respiratory syndrome) coronavirus [49]. Other viruses depend on phosphatidylcholine synthesis, fatty acid synthesis or geranylation. The highly curved nature of cubic membranes, possibly induced by multiple interactions between replicase proteins, might assist virus budding. Virus replication often slows secretion, however, which means that viruses can interfere with membrane trafficking proteins and their regulation by small GTPases [2, 47]. Additional host factors involved in virus-induced membrane remodelling include the SNARE (soluble NSF attachment protein receptor) proteins, which are mediators of vesicle fusion used by the human cytomegalovirus (HCMV), the VAMP-associated proteins (VAP) that bind to SNARE and are used by hepatitis C virus, and the ESCRT (endosomal sorting complex required for transport), which are essential for HIV assembly in the plasma membrane [19, 47, 50].

Inside membranous webs, virus assembly is precisely coordinated with transport of replicated viral genomes from the RC; the specific machinery remains to be identified. Lessons from plant viruses and their movement proteins involved in viral genome transport inside cells and between cells through plasmodesmata can provide some clues for the study of animal viruses [51].

Finally, signalling related to cytoskeleton-mediated organelle movement, particularly in the case of mitochondria, can be essential for factory assembly and virus morphogenesis. Mitochondria attach to membranous webs and are thought to provide energy for viral factory activities. They can have additional roles, as in infected cells it was observed that some mitochondrial proteins abandon the organelle and are integrated into the factories, where they interact with viral proteins [52, 53].

14.5 Perspectives and Conclusions

Some of the following new methodologies have not yet been applied to the study of viral morphogenesis, but their capacity to track macromolecules in cells suggests great potential for the study of virus assembly.

14.5.1 Super-Resolution Fluorescence Microscopy

Limited by the diffraction barrier, conventional light microscopy methods reach resolutions of ~100–500 nm. Many viruses and cell substructures are smaller than this and cannot be solved by LM. In recent years, several laboratories have developed a group of optical microscopy methods, termed super-resolution microscopy, that have improved spatial resolution by an order of magnitude over the diffraction limit [54]. Future improvements in microscopes, fluorescent probes and labelling chemistry will further refine the resolution of these methods, considerably narrowing the gap between light and electron microscopy. Using a variety of technical strategies, these new technologies have begun to provide insights into cell biology and virology. In particular, real-time imaging methods that track individual virus particles in living cells are being used to study virus assembly, and super-resolution microscopy has already defined interactions with cell factors during HIV-1 budding [55, 56]. In the near future, these microscopies will have an increasing impact in the field of virus morphogenesis.

14.5.2 Correlative Microscopy: From Live Cells to High Resolution

Electron microscopy has contributed more than any other method to our understanding of virus assembly in the cell, although its static nature nonetheless makes it difficult to characterise highly dynamic processes. Correlative light and electron microscopy (CLEM) combines the advantages of live cell imaging with the high resolution of EM. A number of procedures have been reported, and the method of choice depends on the research question [37]. Basically, with CLEM we can select individual live cells with interesting features for a detailed, high-resolution study in TEM and ET. Finding adequate probes for CLEM will be the main technical challenge of these studies.

14.5.3 Clonable Tags for Electron Microscopy and Tomography

Genetically clonable tags for TEM and ET would supply new strategies for the ultrastructural characterisation of virus assembly. To date, two types of approaches have been reported, using photoconversion of fluorescent tags or metal-binding proteins. Diaminobenzidine (DAB) can be photoconverted by production of singlet oxygen from fluorescing proteins to generate dense osmiophilic precipitates that are visible in the electron microscope [57]. The resulting signals are diffuse and lack the resolution of particulate probes, although reasonable results have been obtained for proteins concentrated in cell organelles and in electron tomography [37, 42].

Metal-binding proteins were recently validated as clonable tags for EM in bacteria [58, 59] and mammalian cells [60]. These methods allow detection of intracellular proteins at high sensitivity and molecular scale resolution. To track protein molecules in CLEM, proteins can be double-tagged with a metal-binding peptide and a fluorescent probe. This will give us a new way to look inside cells and visualise where and how individual macromolecules come together to build viral particles.

14.5.4 Conclusions

New technologies have recently begun to offer access to analysis of viruses in cells in unprecedented detail. The complexity of the interaction networks established in these contexts is changing our concept of viruses from that of inert molecular organisms to "live" entities able to carry out a wide variety of activities inside cells.

The different technologies and their integration for the study of virus-cell interactions are summarised in Fig. 14.8. From the conventional techniques used to characterise infection kinetics to the new developments in proteomics, genomics, bioinformatics and microscopy, research in this field is generating a vast amount of information about how viruses manage to assemble all the machinery needed to build infectious particles from viral macromolecules and cell materials. Relevant examples on what is being learned on the participation of host cell factors in virus assembly have been described in this chapter by focusing on studies with a few representative model systems. Future work must be accompanied by studies of how cells position their proteins and regulate organelle shape, size and movement. By understanding how viruses manipulate these processes, we will not only learn about viruses but also about cell architecture and compartmentalisation of functions.

Acknowledgements Our gratitude to Drs German Andrés, Abraham Minsky, Ralf Bartenschlager, Petr Chlanda, Jacomine Krijnse-Locker, Laura Sanz, José Jesús Fernández, Juan Fontana, and Noelia López Montero for providing images, and to Catherine Mark for editorial assistance. Work in C.R.'s laboratory was funded by a grant from the Ministry of Science and Innovation of Spain (BIO2009-07255). I.F.C. is recipient of a fellowship from the FPI Program of the Spanish Ministry of Economy and Competitiveness.

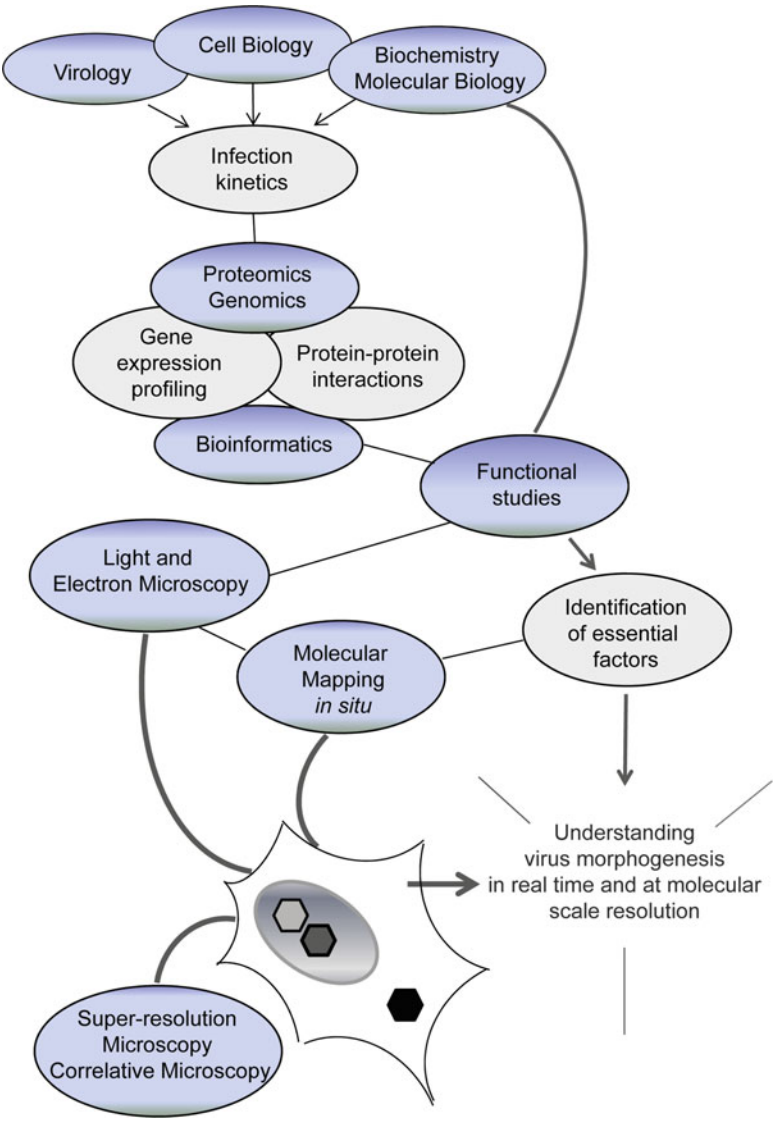


Fig. 14.8 Various technologies and their integration in the study of virus-cell interactions during viral morphogenesis. See text for a description

References and Further Reading

1. Novoa RR, Calderita G, Arranz R, Fontana J, Granzow H, Risco C (2005) Virus factories: associations of cell organelles for viral replication and morphogenesis. Biol Cell 97:147–172
2. Netherton CL, Wileman T (2011) Virus factories, double membrane vesicles and viroplasm generated in animal cells. Curr Opin Virol 1:381–387

3. Nagy PD, Pogany J (2010) Global genomics and proteomics to identify host factors as targets to induce resistance against *tomato bushy stunt virus*. Adv Virus Res 76:123–177
4. Alkhalil A, Hammamieh R, Hardick J, Ichou MA, Jett M, Ibrahim S (2010) Gene expression profiling of monkeypox virus-infected cells reveals novel interfaces for host-virus interactions. Virol J 7:173
5. Friedel CC, Haas J (2011) Virus-host interactomes and global models of virus-infected cells. Trends Microbiol 19:501–508
6. Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershey CL, Rodgers MA, Ramirez JL, Dimopoulos G, Yang PL, Pearson JL, Garcia-Blanco MA (2009) Discovery of insect and human dengue virus host factors. Nature 458:1047–1050
7. Meliopoulos VA, Andersen LE, Birrer KF, Simpson KJ, Lowenthal JW, Bean AG, Stambas J, Stewart CR, Tompkins SM, van Beusechem VW, Fraser I, Mhlanga M, Barichievy S, Smith Q, Leake D, Karpilow J, Buck A, Jona G, Tripp RA (2012) Host gene targets for novel influenza therapies elucidated by high-throughput RNA interference screens. FASEB J 26:1372–1386
8. Pache L, König R, Chanda SK (2010) Identifying HIV-1 host cell factors by genome-scale RNAi screening. Methods 53:3–12
9. Morita E, Sandrin V, McCullough J, Katsuyama A, Hamilton IB, Sundquist WI (2011) ESCRT-III protein requirements for HIV-1 budding. Cell Host Microbe 9:235–242
10. MacFarlane SA, Uhrig JF (2008) Yeast two-hybrid assay to identify host-virus interactions. Methods Mol Biol 451:649–672
11. De Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugué S, Meiffren G, Pradezynski F, Faria BF, Chantier T, Le Breton M, Pellet J, Davoust N, Mangeot PE, Chaboud A, Penin F, Jacob Y, Vidalain PO, Vidal M, André P, Raourdin-Combe C, Lotteau V (2008) Hepatitis C virus infection protein network. Mol Syst Biol 4:230
12. Zhang L, Villa NY, Rahman MM, Smallwood S, Shattuck D, Neff C, Dufford M, Lanchbury JS, Labaer J, McFadden G (2009) Analysis of vaccinia virus-host protein-protein interactions: validations of yeast two-hybrid screenings. J Proteome Res 8:4311–4318
13. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Ireland JT, Chiang CY, Tu BP, De Jesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell 135:49–60
14. Zhang B, Kraemer B, Sengupta D, Fields S, Wickens M (1999) Yeast three-hybrid system to detect and analyze interactions between RNA and protein. Methods Enzymol 306:93–113
15. Lee E-G, Linial ML (2004) Basic residues of the retroviral nucleocapsid play different roles in Gag-Gag and Gag-Z RNA interactions. J Virol 78:8486–8495
16. Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, D'Orso I, Fernandes J, Jan G, Frankel AD, Alber T, Zhou Q, Krogan NJ (2011) Purification and characterization of HIV-human protein complexes. Methods 53:13–19
17. Zhang F, Zang T, Wilson SJ, Johnson MC, Bieniasz PD (2011) Clathrin facilitates the morphogenesis of retrovirus particles. PLoS Pathog 7:e1002119
18. Bartee E, McCormack A, Früh K (2006) Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. PLoS Pathog 2:e107
19. Martin-Serrano J, Neil SJD (2011) Host factors involved in retroviral budding and release. Nat Rev Microbiol 9:519–531
20. Kuhl BJ, Cheng V, Wainberg MA, Liang C (2011) Tetherin and its viral antagonists. J Neuroimmune Pharmacol 6:188–201
21. Gonzalez O, Zimmer R (2011) Contextual analysis of RNAi-based functional screens using interaction networks. Bioinformatics 27:2707–2713
22. Fontana J, López-Montero N, Elliott RM, Fernández JJ, Risco C (2008) The unique architecture of bunyamwera virus factories around the Golgi complex. Cell Microbiol 10:2012–2028

23. Katsafanas GC, Moss B (2007) Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. Cell Host Microbe 2:221–228
24. Sivaraman D, Biswas P, Cella LN, Yates MV, Chen W (2011) Detecting RNA viruses in living mammalian cells by fluorescence microscopy. Trends Biotech 29:307–313
25. Bozzola JJ, Russell LD (1999) Electron microscopy. Principles and techniques for biologists. Jones and Bartlett Publishers, Inc, Boston
26. Goldsmith CS, Miller SE (2009) Modern uses of electron microscopy for detection of viruses. Clin Microbiol Rev 22:552–563
27. Fiala JC (2005) Reconstruct: a free editor for serial section microscopy. J Microsc 218:52–61
28. Schatten H (2011) Low-voltage high-resolution SEM (LVHRSEM) for biological structural and molecular analysis. Micron 42:175–185
29. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, Sun S, Minsky A (2008) Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga* mimivirus. PLoS Biol 6:e114
30. Risco C, Rodríguez JR, López-Iglesias C, Carrascosa JL, Esteban M, Rodríguez D (2002) Endoplasmic reticulum-Golgi intermediate compartment membranes and vimentin filaments participate in vaccinia virus assembly. J Virol 76:1839–1855
31. Briggs JAG, Kräusslich H-G (2011) The molecular architecture of HIV. J Mol Biol 410:491–500
32. Welsch S, Miller S, Romero-Brey I, Merz A, Bleck CKE, Walther P, Fuller SD, Anthony C, Krijnse-Locker J, Bartenschlager R (2009) Composition and three-dimensional architecture of the dengue virus replication and assembly sites. Cell Host Microbe 5:365–375
33. Chlanda P, Carbajal MA, Cyrklaff M, Griffiths G, Krijnse-Locker J (2009) Membrane rupture generates single open membrane sheets during vaccinia virus assembly. Cell Host Microbe 6:81–90
34. Cyrklaff M, Risco C, Fernández JJ, Jiménez MV, Esteban M, Baumeister W, Carrascosa JL (2005) Cryo-electron tomography of vaccinia virus. Proc Natl Acad Sci USA 102:2772–2777
35. Kiernan JA (2008) Histological and histochemical methods. Theory and practice. Scion Publishing Ltd, Oxfordshire
36. Hayat MA (2002) Microscopy, immunohistochemistry, and antigen retrieval methods. Kluwer Academic//Plenum Publishers, New York
37. Van Weering JRT, Brown E, Sharp TH, Mantell J, Cullen PJ, Verkade P (2010) Intracellular membrane traffic at high resolution. Methods Cell Biol 96:619–648
38. Tokuyasu KT (1973) A technique for ultracryotomy of cell suspensions and tissues. J Cell Biol 57:551–565
39. Slot JW, Geuze HJ (2007) Cryosectioning and immunolabeling. Nat Protoc 2:2480–2491
40. Sodeik B, Doms RW, Ericsson M, Hiller G, Machamer CE, van't Hof W, van Meer G, Moss B, Griffiths G (1993) Assembly of vaccinia virus: role of the intermediate compartment between the endoplasmic reticulum and the Golgi stacks. J Cell Biol 121:521–541
41. Jouvenet N, Simon SM, Bieniasz PD (2011) Visualizing HIV-1 assembly. J Mol Biol 410:501–511
42. Shu X, Lev-Ram V, Deerinck TJ, Qi Y, Ramko EB, Davidson MW, Jin Y, Ellisman MH, Tsien RY (2011) A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues and organisms. PLoS Biol 9:e1001041
43. Gaietta G, Deerinck TJ, Admas SR, Bouwer J, Tour O, Laird DW, Sosinsky GE, Tsien RY, Ellisman MH (2002) Multicolor and electron microscopic imaging of connexin trafficking. Science 296:503–507
44. La Scola B, Desnues C, Pagnier I, Robert C, Barrasi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoul D (2008) The virophage as a unique parasite of the giant mimivirus. Nature 455:100–104
45. Carvalho ZG, Alves de Matos AP, Rodrigues-Pousada C (1988) Association of African swine fever virus with the cytoskeleton. Virus Res 11:175–192

46. Rojo G, Chamorro M, Salas ML, Viñuela E, Cuezva JM, Salas J (1998) Migration of mitochondria to viral assembly sites in African Swine fever virus-infected cells. J Virol 72:7583–7588
47. Pierini R, Cottam E, Roberts R, Wileman T (2009) Modulation of membrane traffic between endoplasmic reticulum, ERGIC and Golgi to generate compartments for the replication of bacteria and viruses. Sem Cell Dev Biol 20:828–833
48. Fontana J, Tzeng WP, Calderita G, Fraile-Ramos A, Frey TK, Risco C (2007) Novel replication complex architecture in rubella replicon-transfected cells. Cell Microbiol 9:875–890
49. Deng Y, Almshwerqi ZA, Ng MML, Kohlwein SD (2010) Do viruses subvert cholesterol homeostasis to induce host cubic membranes? Trends Cell Biol 20:371–379
50. Liu STH, Sharon-Friling RS, Ivanova P, Milne SB, Myers DS, Rabinowitz JD, Brown HA, Shenk T (2011) Synaptic vesicle-like lipidome of human cytomegalovirus virions reveals a role for SNARE machinery in virion egress. Proc Natl Acad Sci USA 108:12869–12874
51. Harries P, Ding B (2011) Cellular factors in plant virus movement: at the leading edge of macromolecular trafficking in plants. Virology 411:237–243
52. Fontana J, López-Iglesias C, Tzeng WP, Frey TK, Fernández JJ, Risco C (2010) Three-dimensional structure of rubella virus factories. Virology 405:579–591
53. Ilkow CS, Weckbecker D, Cho WJ, Meier S, Beatch MD, Goping IS, Herrmann JM, Hobman TC (2010) The rubella virus capsid protein inhibits mitochondrial import. J Virol 84:119–130
54. Huang B, Babcock H, Zhuang X (2011) Breaking the diffraction barrier: super-resolution imaging of cells. Cell 143:1047–1058
55. Jouvenet N, Zhadina M, Bieniasz PD, Simon SM (2011) Dynamics of ESCRT protein recruitment during retroviral assembly. Nat Cell Biol 13:394–401
56. Lehmann M, Rocha S, Mangeat B, Blanchet F, Uji-I H, Hofkens J, Piguet V (2011) Quantitative multicolor super-resolution microscopy reveals tetherin HIV-1 interaction. PLoS Pathog 7: e1002456
57. Monosov EZ, Wenzel TJ, Luers GH, Heyman JA, Subramani S (1996) Labeling of peroxisomes with green fluorescent protein in living *P. pastoris* cells. J Histochem Cytochem 44:581–589
58. Diestra E, Fontana J, Guichard P, Marco S, Risco C (2009) Visualization of proteins in intact cells with a clonable tag for electron microscopy. J Struct Biol 165:157–168
59. Wang Q, Mercogliano CP, Löwe J (2011) A ferritin-based label for cellular electron cryotomography. Structure 19:147–154
60. Risco C, Sanmartín-Conesa E, Tzeng WP, Frey TK, Seybold V, de Groot RJ (2012) Specific, sensitive, high-resolution detection of protein molecules in eukaryotic cells using metal-tagging transmission electron microscopy. Structure 20:759–766

Further Reading

Baumgärtel V, Ivanchenko S, Dupont A, Sergeev M, Wiseman PW, Kräusslich HG, Bräuchle C, Müller B, Lamb DC (2011) Live-cell visualization of dynamics of HIV budding site interactions with an ESCRT component. Nat Cell Biol 13:469–474

Claverie JM, Abergel C (2009) Mimivirus and its virophage. Annu Rev Genet 43:49–66

Erickson KD, Bouchet-Marquis C, Heiser K, Szomolanyi-Tsuda E, Mishra R, Lamothe B, Hoenger A, Garcea RL (2012) Virion assembly factories in the nucleus of polyomavirus-infected cells. PLoS Pathog 8:e100263

Fogarty KH, Zhang W, Grigsby IF, Johnson JL, Chen Y, Mueller JD, Mansky LM (2011) New insights into HTLV-1 particle structure, assembly and Gag-Gag interactions in living cells. Viruses 3:770–793

Fu C, Johnson J (2011) Viral life cycles captured in three-dimensions with electron microscopy tomography. Curr Opin Virol 1:125–133

Iwasaki K, Omura T (2010) Electron tomography of the supramolecular structure of virus-infected cells. Curr Opin Struct Biol 20:632–639

Mutsafi Y, Zauberman N, Sabanay O, Minsky A (2010) Vaccinia-like cytoplasmic replication of the giant mimivirus. Proc Natl Acad Sci USA 107:5978–5982

Nagy PD, Pogany J (2012) The dependence of viral RNA replication on co-opted host factors. Nat Rev Microbiol 10:137–149

Otto A, Bernhardt J, Hecker M, Becher D (2012) Global relative and absolute quantitation in microbial proteomics. Curr Opin Microbiol 15:1–9

Rust M, Lakadamyali M, Brandenburg B, Zhuang X (2008) Single-virus tracking in live cells. In: Selvin PR, Ha T (eds) Single molecule techniques: a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Van Vliet K, Mohamed MR, Zhang L, Villa NY, Werden SJ, Liu J, McFadden G (2009) Poxvirus proteomics and virus-host protein interactions. Microbiol Mol Biol Rev 73:730–749

Also especially recommended for further reading are references [1, 2, 3, 5, 19, 37, 49, 54] *listed above*.

Chapter 15
Virus-Receptor Interactions and Receptor-Mediated Virus Entry into Host Cells

José M. Casasnovas

Abstract The virus particles described in previous chapters are vehicles that transmit the viral genome and the infection from cell to cell. To initiate the infective cycle, the viral genome must therefore translocate from the viral particle to the cytoplasm. *Via* distinct proteins or motifs in their outermost shell, the particles attach initially to specific molecules on the host cell surface. These virus receptors thus mediate penetration of the viral genome inside the cell, where the intracellular infective cycle starts. The presence of these receptors on the cell surface is a principal determinant of virus host tropism. Viruses can use diverse types of molecules to attach to and enter into cells. In addition, virus-receptor recognition can evolve over the course of an infection, and virus variants with distinct receptor-binding specificities and tropism can appear. The identification of virus receptors and the characterization of virus-receptor interactions have been major research goals in virology for the last two decades. In this chapter, we will describe, from a structural perspective, several virus-receptor interactions and the active role of receptor molecules in virus entry.

Keywords Virus-host • Virus tropism • Virus attachment • Virus-receptor • Virus structure • Virus entry • Virus neutralization • Capsid dynamics • Uncoating • Membrane penetration • Endocytosis • Cell surface • Cell adhesion • Cell surface molecules • Membrane proteins • Glycoproteins • Carbohydrates • Protein interactions • Molecular recognition • Crystallography • Cryo-EM

J.M. Casasnovas (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: jcasasnovas@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_15,

Abbreviations

Ad	Adenovirus
APN	Aminopeptidase N
CAR	Coxsackievirus-adenovirus receptor
cryo-EM	Cryo-electron microscopy
CV	Coxsackievirus
D	Domain
DAF	Decay-accelerating factor
DC	Dendritic cells
DC-SIGN	(DC-specific ICAM-3-grabbing nonintegrin)
EFN	Ephrin
EV	Echovirus
FMDV	Foot-and-mouth disease virus
g/gp	Glycoprotein
H	Hemagglutinin
HA	Influenza A hemagglutinin
HeV	Hendra virus
HIV-1	Human immunodeficiency virus-type 1
HN	Hemagglutinin neuraminidase
HRV	Human rhinovirus
HS	Heparan sulphate
HSV	Herpes simplex virus
ICAM-1	Intercellular adhesion molecule-1
IgSF	Immunoglobulin superfamily
LDLR	Low-density lipoprotein receptor
MCP	Membrane cofactor protein
MV	Measles virus
N	Neuraminidase
NAG	N-acetyl-glucosamine
NDV	Newcastle disease virus
NiV	Nipah virus
PM	Plasma membrane
PV	Poliovirus
PVR	Poliovirus receptor
SCR	Short consensus repeats
SLAM	Signalling lymphocytic activation molecule
VP	Viral capsid protein.

15.1 Introduction: Virus Entry into Host Cells, the Recognition of Cell Surface Molecules

The viral particles formed in infected host cells (see Chaps. 10, 11, 12, 13, and 14) are metastable structures that transmit the viral genome and the infection from cell to cell. Viruses must therefore penetrate host cells to initiate the replicative infective cycle by exploiting the cell machinery. In the extracellular transit stage of the viral cycle, animal viruses and bacteriophages attach to specific cell surface molecules (virus receptors) suited for host cell entry following virus-specific entry pathways. Virus receptors must be distinguished from attachment factors, surface molecules to which some viruses can bind but that do not themselves promote virus entry into host cells [1]. Virus binding to attachment factors concentrates virus particles onto the cell surface, which can help viruses to encounter specific entry receptors that mediate genome translocation into the cytoplasm. The virus receptor molecules are not just required for initial virus binding to host cells, but also for the transfer of the viral genome through cellular membranes [1]. Virus-receptor interactions can trigger changes in the virus particles that initiate genome translocation, or alterations in the cell, such as signalling events that facilitate virus entry [2]. Viruses enter host cells at the cell *surface* or after endocytosis (Fig. 15.1). Multivalent binding of the virus particles to receptors on the cell surface can mediate uncoating or release of the viral genome in non-enveloped viruses and/or fusion of the virus and cell membranes (Fig. 15.1). Moreover, viruses bound to cell surface molecules can be internalized by following different endocytic pathways [3], where exposure to low pH, enzymatic modification or other cellular factors leads to the delivery of the genome into the cytoplasm (Fig. 15.1).

Viruses evolve to recognize specific cell surface receptor molecules appropriate for productive entry and infection of host cells, which frequently determines the host tropism or the cell type a virus can infect. Selection of cell entry receptor by viruses appears to be determined by subtle interactions that regulate the specificity and affinity necessary for efficient cell attachment. Virus-receptor interaction can nonetheless be a highly dynamic process. A single virus can recognize one or several cell entry molecules, which can also differ among virus variants or during the course of an infection [4, 5]. Virus recognition of receptors is under continuous evolutionary pressure to increase their infection efficiency, which can lead to the emergence of virus variants with altered infectivity or tissue tropism.

In this chapter, several examples of animal virus-receptor interactions will be presented, together with a description of known models of receptor-switching viruses. The chapter presents a structural view of some virus-receptor interactions that have been characterized by structures of complexes. We will also describe how the viral genome exits the capsid (uncoating) in some non-enveloped animal viruses, illustrating the role of cell surface receptor molecules in the entry process. Membrane penetration events in enveloped viruses are discussed in Chap. 16. Specific aspects of receptor recognition and injection-mediated genome uncoating by some bacteriophages are described in Chap. 17.

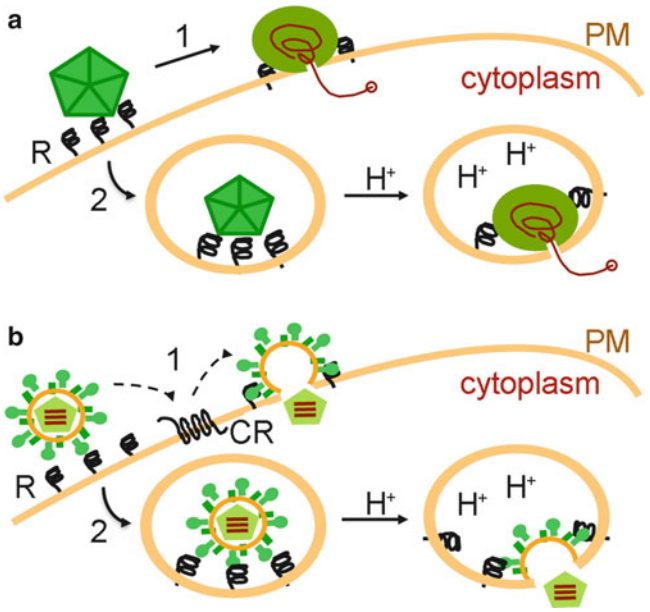


Fig. 15.1 Virus entry into host cells. Scheme illustrating two cell entry pathways by non-enveloped (**a**) or enveloped viruses (**b**). pH-independent entry at the cell surface (*pathway 1*): Genome uncoating and penetration into cytoplasm (**a**) or virus-cell membrane fusion (**b**) at the plasma membrane (PM). Virus binding to a single receptor (R) or to additional co-receptor molecules (CR) at neutral pH triggers genome translocation. pH-dependent entry in endosomes (*pathway 2*): Receptor-mediated endocytosis of viruses prior to penetration into the cytoplasm. In the endosomes, viruses firmly bound to multiple receptor molecules are exposed to progressively lower pH, which triggers genome uncoating and penetration (**a**) or fusion (**b**). The entry reaction is mediated by acidification of endosomal compartments, with or without receptor priming

15.2 Virus-Receptor Interactions and Receptor Specificity Switch

The known number and diversity of cell surface molecules exploited by viruses to enter host cells is still increasing [6]. Viruses recognize a variety of cell surface molecules specifically, including glycolipids, carbohydrates and proteins with very distinct folding structures. There are viruses specific for a single receptor molecule, whereas others bind to several structurally distinct receptors for host cell entry [2]. The virus particles use certain proteins in their outermost shell to attach to the cell surface molecules. The multivalent nature of the particles allows viruses to bind with very high avidity to the host cell surface, even though monomeric virus-receptor interactions are usually of moderate affinity (0.1–1 μM) [7].

Some non-enveloped viruses have proteins specialized in attachment to cell surface receptors; however, in other cases unique motifs on the naked capsids are engaged in receptor recognition. Here we discuss both receptor recognition modes,

providing a description of adenovirus and picornavirus binding to cell surface molecules. The enveloped viruses bear membrane-bound glycoproteins that bind to the receptor and trigger virus-cell membrane fusion during the entry process, as will be described in Chap. 16. We present the diverse types of cell surface receptors recognized by human immunodeficiency virus and paramyxoviruses, which have been characterized. In the last part of this section, we illustrate the importance of cell surface carbohydrates in virus attachment and infection of host cells.

15.2.1 Virus-Receptor Interactions in Non-Enveloped Viruses

In this subsection, we describe receptor recognition by two families of non-enveloped viruses (picornaviruses and adenoviruses) for which virus-receptor interactions have been characterised in detail by structural studies. Picornaviruses use their naked capsid to attach to receptors, whereas adenoviruses have fibres that protrude from the capsid for attachment to several cell surface molecules.

Picornaviruses and Their Receptors

The picornaviruses are a large family of non-enveloped viruses responsible for numerous human and animal diseases. Picornavirus particles are formed by an icosahedral protein capsid built by three external viral capsid proteins (VP1 to VP3) and the internal protein VP4 packed inside with a single-stranded RNA genome (Fig. 15.2a). The capsid is composed of 60 basic subunits or protomers arranged as 12 pentamers (see Chap. 10). The members of this virus family bind to distinct types of receptor molecules suited for entry into host cells [8].

The poliovirus receptor (PVR) was one of the first picornavirus receptors to be characterized [9]. PVR is a type I membrane protein and a member of the immunoglobulin superfamily (IgSF), which has three Ig-like domains (D1 to D3) at the extracellular region (Fig. 15.2a). The receptor for most (90 %) identified human rhinovirus (HRV) serotypes, the major group of HRV, is intercellular adhesion molecule-1 (ICAM-1) [10, 11], another IgSF member with five Ig-like extracellular domains (Fig. 15.2a). PV and HRV bind similarly to the N-terminal membrane distal domains of the receptor molecules (Fig. 15.2b) [8]. Both viruses use a depressed surface, or canyon, formed by two neighbouring protomers around the five-fold icosahedral vertices of the capsid (Figs. 15.2a, b). Cryo-electron microscopy (cryo-EM; see Chap. 3) and binding studies show some differences in the way PV and HRV bind to their IgSF receptors. Kinetics for monomeric receptor binding to the virus particles showed more rapid kinetic binding rates to PV than to HRV [12]. Moreover, cryo-EM structures of virus-receptor complexes showed that PV uses more exposed residues on the canyon walls than HRV [8, 12].

The use of the canyon for binding to cell surface receptors was also described for other picornaviruses such as the coxsackieviruses (CV) A21 and B3, and echovirus 1 (EV1) [8, 13]. CVA21 binds to ICAM-1, CVB3 to the coxsackievirus-adenovirus receptor (CAR), and echovirus 1 (EV1) to the $\alpha_2\beta_1$ integrin. The CAR protein also

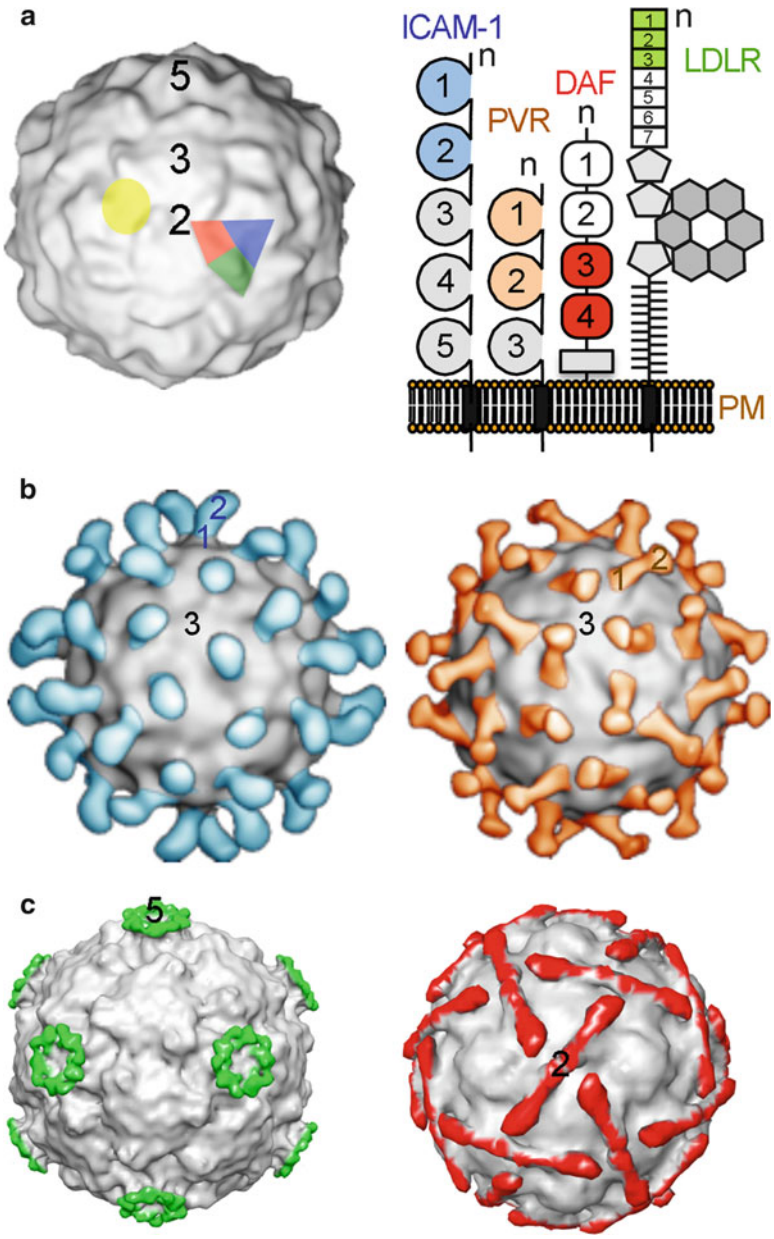


Fig. 15.2 Virus-receptor interactions in picornaviruses. (**a**) *Left*: Surface representation of the poliovirus (PV) particle computed from the crystal structure of PV1/M (PDB ID 1ASJ). Numbers indicate location of the icosahedral two-, three- and five-fold axes. The *triangle* illustrates the approximate location of a capsid building block or protomer, assembled from the external VP1 (*blue*), VP2 (*green*) and VP3 (*red*). The canyon region is marked with a *yellow circle*. *Right*: Scheme of picornavirus receptors ICAM-1, PVR, DAF and LDLR. The N-terminus (n) and

belongs to the IgSF, but the $\alpha_2\beta_1$ integrin is a non-IgSF member that fits into the picornavirus canyon. Cryo-EM structural studies showed that the I-domain of the integrin penetrates the EV1 capsid canyon [13], as described for PV, HRV and CVA21 binding to IgSF receptor molecules. Thus, the canyon is a well-defined receptor-binding region in picornaviruses [8]. The concave nature of the canyon is thought to be suited for hiding conserved receptor-binding residues from neutralizing antibodies, which cannot penetrate the depressed surface as efficiently as relatively narrower receptor molecules. This hypothesis was nonetheless challenged by studies showing an antibody that penetrates the canyon [14], although it interacts mostly with the walls rather than the bottom of the canyon. Receptor-binding residues in PV are more exposed on the canyon walls and can thus be targeted by antibodies elicited by PV vaccines [8, 12]. The use of recessed virus surfaces in receptor recognition is relatively common in virus-receptor interactions, and could be a viral strategy to protect some receptor-binding residues from immune surveillance and/or might have arisen to increase surface contact area and binding energy for cell receptor molecules.

Other picornaviruses do not use the canyon for attachment to cell entry receptors; viruses of this family, such as foot-and-mouth disease virus (FMDV), some HRV and echoviruses (EV), use exposed regions of the capsid. FMDV exploits an RGD motif exposed in the VP1 GH loop to bind to integrin receptors [15]. This receptor-binding motif is accessible and is a major antigenic site; however, virus can escape from immune neutralization by mutating some residues adjacent to the RGD, which prevents detection by some antibodies while preserving cell attachment activity and viability. Approximately 10 % of HRV serotypes, the minor group, do not bind to ICAM-1, and use members of the low density lipoprotein receptor (LDLR) family to enter host cells (Fig. 15.2a) [16]. Even though minor and major group HRV are closely related, the mode by which they bind their respective receptors is strikingly different (Figs. 15.2b, c). Minor group HRV do not use the canyon, and contact the receptor through a protruding region close to the capsid five-fold axis (Fig. 15.2c, left) [17]. These HRV bear a conserved Lys residue in VP1 that contacts an acidic cluster and a Trp residue in the N-terminal ligand binding domains of the LDLR proteins. This lysine is absent in major group HRV that bind to ICAM-1, and is therefore likely to be a major determinant of the distinct receptor-binding specificity described for rhinoviruses.

Fig. 15.2 (continued) extracellular domains are labelled. Receptor domains used to determine structures of virus-receptor complexes shown below are coloured. (**b**) Cryo-EM structure of HRV3-ICAM-1 (*left*) and PV1-PVR (*right*), picornaviruses that use the canyon for receptor binding. The complexes were prepared in solution with purified virus particles and two-domain (1 and 2) fragments of the receptor molecules. Location of the domains is shown. D1 penetrates the canyon, whereas D2 does not contact the virus and protrudes from the capsid. Images provided by Holland Cheng and Li Xing, adapted from references [12, 79]. (**c**) Cryo-EM structure of HRV2-LDLR (EMD-1049) and EV12-DAF (EMD-1057) complexes, representative of picornaviruses that do not use the canyon for receptor binding. Ligand binding repeats 1–3 of the very low-density lipoprotein receptor or domains 3–4 of DAF were used to prepare complexes. Surfaces of bound receptors are coloured as in (**a**). Images prepared with Chimera (cgl.ucsf.edu/chimera) from cryo-EM maps

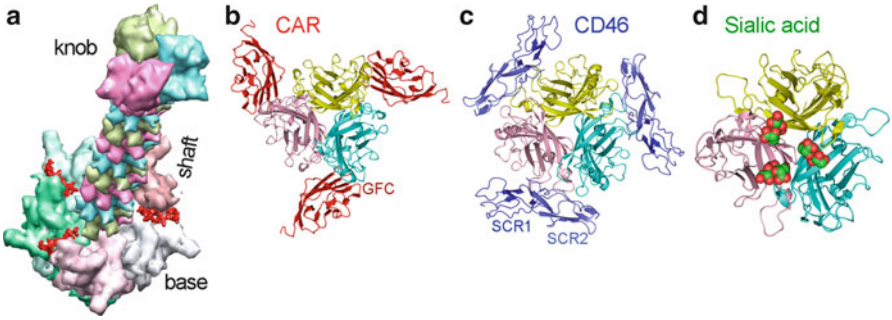


Fig. 15.3 Virus-receptor interactions in adenoviruses. Representation of the receptor-binding adenovirus fibre and penton base (**a**) (see also Chap. 11), and crystal structures of the trimeric Ad knob in complex with CAR (**b**), CD46 (**c**), or sialic acid (**d**). Ribbon drawings are shown for the complexes prepared with the N-terminal (D1) of CAR (*red*) and Ad12 (PDB ID 1KAC), or with the two N-terminal short consensus repeats (SCR1, 2) of CD46 (*blue*) and Ad11 (PDB ID 2O39). The Ad-binding GFC β-sheet of the CAR domain is indicated. Sialic acids in complex with the Ad37 knob (PDB ID 1UXA) are shown with carbons (*green*) and oxygens (*red*) as spheres. Figure (a) was provided by Carmen San Martín and figures (b–d) were prepared with PyMOL (pymol.org)

A picornavirus-receptor interaction different from those described above was reported for EV using decay-accelerating factor (DAF, CD55) as a cell entry receptor [18]. These viruses bind to DAF through a capsid protrusion at the southern rim of the canyon, distant from the five-fold axis and near the capsid two-fold axis (Fig. 15.2c, right) [19, 20]. The receptor-binding region in EV11 protrudes and is more exposed than the canyon. The kinetic association rate for the monomeric EV11-CD55 interaction is thus one to two orders of magnitude higher than those shown for picornaviruses that use the canyon for receptor recognition, which probably reflects the distinct nature of the virus-receptor interactions [21].

The structural studies of picornavirus-receptor complexes discussed here show that viruses of this family can use diverse capsid regions for recognition of cell surface molecules and for productive entry into host cells.

Adenoviruses, Non-enveloped Viruses That Bind to Several Cell Surface Molecules

Adenoviruses (Ad) are non-enveloped viruses with icosahedral capsids built by hexon and penton capsomers (see Chap. 2 for definition of capsomers and Chap. 11 for additional information and references). The pentons locate at each of the 12 icosahedral vertices of the capsid, and forms the base of protruding trimeric fibres (Fig. 15.3a). The penton base and the associated fibres form a complex that recognises cell surface receptor molecules [22, 23]. The most capsid-distal, globular region of the fibre, the knob (Fig. 15.3a), mediates initial Ad attachment to the cell, whereas the penton base is used for subsequent tight attachment of the virus to a secondary receptor that mediates virus internalization and host cell entry [22]. Receptor-bound Ad are transported to endosomes, where acidification triggers capsid disassembly and virion penetration into the cytoplasm.

Cell tropism and receptor recognition is well documented for Ad. This group of non-enveloped viruses can use at least three distinct types of receptor molecules for initial attachment to the cell surface through the fibre proteins [23], which is mediated by the most distal trimeric knob structure (Fig. 15.3). Subgroups A, C, D, E and F attach to the CAR receptor, a member of the IgSF found in epithelial tight junctions. CAR has two Ig-like domains at the extracellular region, and mediates homotypic cell-cell interactions through the same face (GFC β-sheet) of the N-terminal domain engaged by the Ad fibres (Fig. 15.3b) [24, 25]. Fibre binding to CAR interferes with homotypic cell-cell adhesion, which destabilizes epithelial cell layers and facilitates virus release to the airway lumen and spreading to a new host [26].

Most subgroup B Ad use the ubiquitous membrane cofactor protein (MCP, CD46) for initial attachment to the host cell [22]. The Ad knob contacts the two N-terminal short consensus repeats (SCR) of the CD46 molecule (Fig. 15.3c) [27]. Although all Ad use the fibre knob for receptor binding, the mode of contact of the knobs with the CAR or the CD46 proteins is distinct (Fig. 15.3). The switch in receptor-binding specificity between CAR and CD46-binding Ad appear to be related to the distinct conformation of the exposed loops in the periphery of the knob. CD46-binding Ad have an extended knob AB-loop that prevents binding to the CAR receptor, as well as specific structural features in the loops that contact the CD46 molecule [5]. Certain Ad can also attach to cell surface carbohydrates, such as heparin or sialic acid [22, 28, 29]. Subgroup D adenoviruses 8, 19 and 37, which recognize sialic acid, are responsible for epidemic keratoconjunctivitis, a very contagious ocular disease. The sialic acid binding site has been mapped also at the fibre knob, but it locates closer to the centre of the trimer than the sites used to bind to the other receptors (Fig. 15.3c) [28, 29]. The Ad-receptor interaction described here illustrate how different Ad evolved to use distinct knob surfaces to attach to different cell surface receptor molecules; this certainly translates in cell tropism and pathogenesis diversity among the member of this virus family.

Following initial fibre attachment to cell surface receptors, the Ad particles engage cell surface integrin molecules using RGD motifs exposed in loops at the capsid penton base, which is a necessary step for host cell entry by endocytosis [30]. The multivalent interaction of the Ad particle with the integrins mediates its clustering, triggering intracellular signals, rearrangement of actin, and clathrin-mediated endocytosis of the virions [22, 23]. During endocytosis, the virion disassembles and subsequently penetrates the membrane. Ad are thus an example of non-enveloped viruses that use two distinct capsid structures to bind different cell surface receptor molecules for attachment to or penetration of host cells.

15.2.2 Virus-Receptor Interactions in Enveloped Viruses

Enveloped viruses bear membrane-bound glycoproteins specialized in the recognition of cell surface molecules and in subsequent fusion of the viral and cell membranes. The receptor-binding and fusion proteins can be one or two distinct

polypeptides, in many cases resulting from the cleavage of a single precursor protein. Both of these proteins associate in the viral envelope, forming the envelope spikes. Among enveloped viruses, the fusion proteins are less diverse than receptor-binding proteins and will be presented in Chap. 16. Here we will discuss the interactions of some viral envelope spikes with cell surface receptor molecules, described by structures of virus-receptor complexes.

The Human Immunodeficiency Virus Cell Attachment Process

Attachment of the human immunodeficiency virus-type 1 (HIV-1) to cell entry receptors and subsequent fusion have been described in great detail. The HIV-1 particle bears two non-covalently bound viral glycoproteins, gp120 and gp41, which associate to form trimeric envelope spikes [31, 32]. There are around 15 spikes per particle [32]. The gp120 glycoprotein mediates attachment of the HIV-1 particles to cell surface receptors, whereas gp41 catalyses fusion of the virus and the cell membrane, and the release of the nucleoprotein to the cell cytoplasm [31, 33] (see Chap. 16). The HIV-1 entry process is relatively complex [33]. gp120 requires engagement of two distinct cell surface molecules to trigger membrane fusion, which occurs at the cell plasma membrane [34]. gp120 initially engages the lymphocyte-specific CD4 [33], an IgSF cell surface protein. CD4 is a type I membrane protein composed of four Ig-like domains in the extracellular region that forms part of the T cell receptor complex in a subset of T lymphocytes. A depressed groove or pocket in gp120 surrounded by hypervariable loops engages the CD4 N-terminal Ig-like domain with high affinity (Fig. 15.4) [35]. A Tyr residue protruding from the CD4 domain penetrates deeply into the gp120 pocket. The CD4-binding site is surrounded by long hypervariable loops and glycans, which further hide the depressed site from immune surveillance [36].

Binding of gp120 to CD4 triggers conformational changes in the HIV-1 protein that expose a conserved structure (bridging sheet) that binds to a second molecule termed co-receptor [35]. Engagement of the co-receptor molecule is essential for virus-cell membrane fusion and for HIV-1 penetration into the host cell. Two distinct HIV-1 co-receptors have been identified, CXCR4 and CCR5, which are chemokine receptors containing seven transmembrane segments [37]; HIV-1 interaction with CCR5 and CXCR4 is strain-specific. HIV-1 strains using CCR5 (R5 viruses) are mainly associated to sexual transmission. They can also evolve to infect T cells by acquiring the ability to use the CXCR4 receptor (X4 viruses). HIV-1 viruses that use both co-receptors are termed R5X4 viruses.

HIV-1 particles are also able to attach specifically to cell surface molecules that do not mediate cell entry and infection, but are used instead to present the virus to CD4-expressing T cells. This process, termed trans-infection, is very efficient and occurs when antigen-presenting dendritic cells (DC) encounter T cells [38]. HIV-1 can therefore bind to DC in the periphery, which transport the virus to lymphoid organs, where it is transferred to T cells that become infected. The C-type lectin DC-SIGN (DC-specific ICAM-3-grabbing nonintegrin) was the first HIV-1 receptor in DC to be characterized as a mediator of trans-infection [39]. DC-SIGN is specific for simple high-mannose glycans and fucose-containing glycosylations

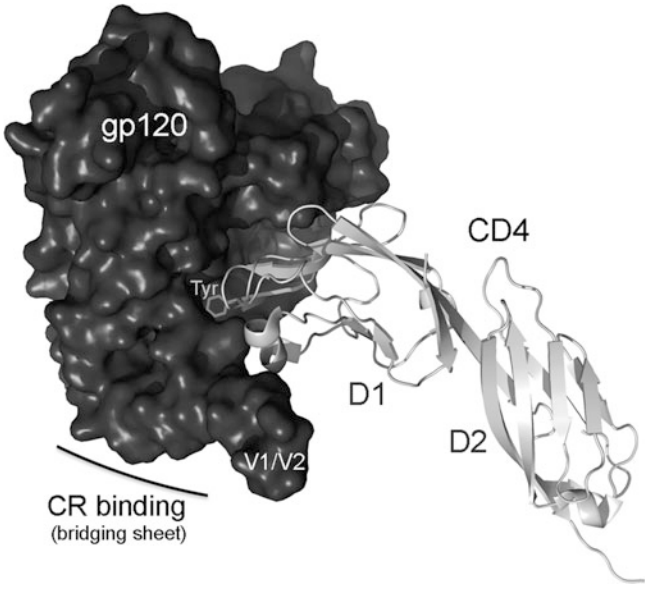


Fig. 15.4 HIV-1 receptor recognition. Crystal structure of the HIV-1 gp120 protein (*dark surface*) bound to the CD4 receptor (*ribbon drawing in grey*) (PDB ID 1GC1). The N-terminal domain of CD4 (D1) fits into a concave gp120 surface, with a deeply penetrating Tyr residue. The location is indicated of the truncated V1/V2 loop and the bridging sheet region that binds to HIV-1 entry co-receptors (CR). CR binding is necessary for virus-cell fusion at the plasma membrane. This and following Figs. 15.5 and 15.6 were prepared with PyMOL (pymol.org)

[40, 41]. gp120 is heavily glycosylated; its glycan composition is mostly of the high-mannose type, which is recognized by DC-SIGN. The glycan shield of the HIV-1 spike is thus used for attachment to DC that trans-infect T cells. It was recently shown that, in addition to the spike glycosylations, HIV-1 also uses charged glycans in its envelope lipidome to attach to DC [42, 43]. The HIV-1 envelope membrane is enriched in certain glycosphingolipids called gangliosides. A specific sialic acid-containing subset of gangliosides appears to recognize receptors on the DC surface [42, 43]. HIV-1 thus exploits different envelope components to attach to distinct cell types, which leads to spread of the infection.

Cell Receptor Recognition in Paramyxovirus

The paramyxoviruses are enveloped, negative-stranded RNA viruses that include serious human and animal pathogens [44]. In viruses of this family, cell attachment and virus-cell membrane fusion are mediated by two distinct membrane glycoproteins (see also Chap. 16). The paramyxovirus attachment proteins are type II membrane proteins anchored to the virus envelope by a single transmembrane domain [44, 45]. Their extracellular region can be divided into an N-terminal stalk region that serves as a spacer and a C-terminal globular domain with receptor binding activity. The C-terminal globular domains of paramyxovirus attachment

glycoproteins all fold into similar six-bladed β-propeller structures [44, 45]. In the virus envelope, the attachment proteins are present as disulphide-linked homodimers, with indications of tetramer formation in some cases [45]. The attachment proteins form complexes with the fusion proteins, which are homotrimers in the viral envelope. Receptor binding triggers rearrangements in these heterocomplexes that alter the structure of fusion proteins, resulting in fusion of viral and cell membranes at neutral pH [45–47] (see Chap. 16).

There is certain diversity in cell surface receptor usage among paramyxoviruses. Rubulaviruses (mumps virus), avulaviruses (Newcastle disease virus) and respiroviruses (Sendai virus) bind to cell surface sialic acids *via* the hemagglutinin neuraminidase (HN) attachment glycoprotein, a bifunctional protein engaged in recognition and hydrolysis of sialic acids. Neither activity is found in the haemagglutinin (H) of morbilliviruses (measles virus), or in the attachment glycoprotein G of henipaviruses (Hendra and Nipah viruses) or pneumoviruses (respiratory syncytial virus) [44]. Measles virus H (MV-H) and the Hendra (HeV-G) and Nipah G (NiV-G) proteins lack the conserved residues that mediate sialic acid binding and hydrolysis [5]. These viruses do not bind to sialic acid carbohydrates, but rather attach to cell surface proteins. HeV and NiV-G interact with the ephrin-B2 or -B3 receptors (EFNB2, EFNB3) [48–50], whereas MV-H can bind CD46 [51, 52], signalling lymphocytic activation molecule (SLAM) [53] or nectin-4 [54, 55], depending on the MV strain. All strains of MV bind to SLAM expressed on macrophages, DC, and lymphocytes, cells where infection starts and develops. MV uses nectin-4 to infect epithelial cells and cross the airway epithelium for transmission to new hosts [54, 55].

Structural studies defined the receptor binding modes of several paramyxoviruses. Both sialic acid and the EFNB2/EFNB3 receptors are recognized by overlapping sites at the recessed center of the β-propeller domain of the paramyxovirus attachment proteins (Fig. 15.5a) [44, 45, 56–58]. Protruding hydrophobic residues at the long GH loop of EFNB2 interact with residues in NiV-G that lie very close to the sialic acid binding site in HN at the central cavity of the β-propeller domain (Fig. 15.5a). The receptor binding mode in measles virus (MV) nonetheless differs from that of other paramyxoviruses. The structures of CD46 and SLAM in complex with MV-H protein show that MV uses the side of the β-propeller domain to bind to cell surface receptor molecules (Figs. 15.5a, b) [59, 60]. The recessed center of the β-propeller, the site of sialic acid binding in several paramyxoviruses, is closed off by a glycan in the MV-H protein. The receptor-binding region in MV-H includes a groove formed by the blades β4 and β5 in the β-propeller, a region with the largest structural difference between MV-H and the other paramyxovirus proteins [59]. Therefore, the receptor-binding regions in paramyxoviruses preserve the recessed nature shared by many receptor-binding sites in virus proteins (Figs. 15.5c, d). In addition, the receptor-binding regions contain a hydrophobic socket that is particularly recessed. This socket accommodates sialic acid in the paramyxovirus HN proteins or receptor-specific features in MV-H (Fig. 15.5e) [56, 59]. In the case of the MV-H, the recessed receptor-binding surface is more extended than in other paramyxovirus attachment protein using the centre of the β-propeller for attachment to cell surface receptors (Figs. 15.5c, d). This MV-H surface can recognise three distinct receptor molecules,

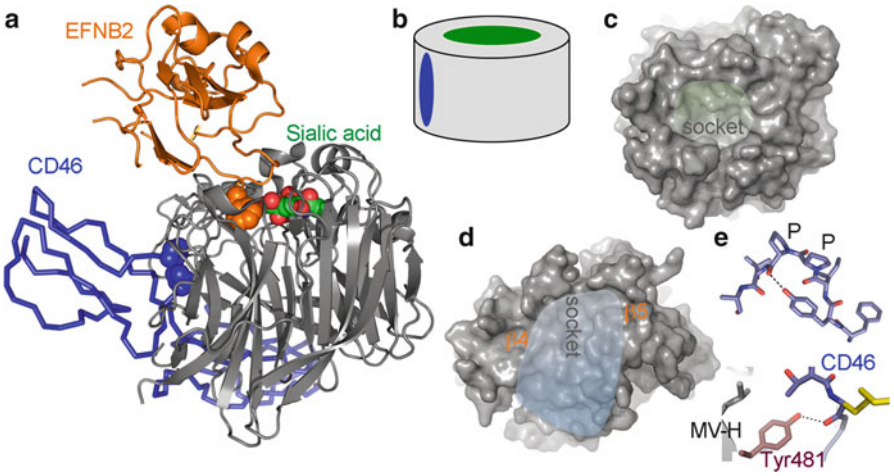


Fig. 15.5 Cell receptor recognition in paramyxovirus. (**a**) Ribbon drawing of the β-propeller domain of a representative paramyxovirus attachment protein (NDV-HN structure; *grey*), bound to three paramyxovirus receptors: Sialic acid (carbons (*green*) and oxygens (*red*) as spheres), the EFNB2 receptor (*orange*) and CD46 (*blue*). Two virus binding Pro residues in CD46 and one Phe in EFNB2 are shown as spheres. The figure was prepared with the crystal structures of complexes of NDV-HN with sialic acid (PDB ID 1E8T), of NiV-G with EFNB2 (PDB ID 2VSM), and of MV-H with CD46 (PDB ID 3INB) after superimposition of the virus proteins. NiV-G and MV-H structures are not shown. (**b**) Cylinder representing a β-propeller domain and location of main receptor binding surfaces in paramyxovirus attachment proteins, based on the complexes shown in (**a**); centre (*green*) and side (*blue*) of the propeller. (**c**) Top view of the NDV-HN structure showing the recessed sialic acid binding surface on the centre of the β-propeller (*light green*), with the socket in which the sialic acid binds (indicated). (**d**) Side view of the MV-H structure with the extended concave surface (*light blue*) use to bind to the SLAM, CD46 and nectin-4 receptors. The β-propeller blades β4 and β5 that form the receptor-binding surface and the socket on the surface are indicated. (**e**) Key interactions for MV recognition of CD46. *Top*. Stick diagram of protruding CD46 loop with two contiguous Pro residues that penetrates the MV-H socket. *Bottom*. Hydrogen bond (*dashed black lines*) of MV-H Tyr481 with a main chain carbonyl of CD46, required for MV binding to CD46 [59]. Oxygens are *red* and nitrogens *dark blue*

CD46, SLAM and nectin-4. As shown for the MV-H/CD46 complex [59], viral residues within the extended receptor-binding groove can show some variability, whereas less accessible residues are conserved; this leads to increasing binding affinity to alternative receptor molecules, which has an important implications on MV tropism and pathogenesis. Single residue mutations in the MV-H groove, such as the Asn481 to Tyr481 switch, is sufficient for MV binding to the ubiquitous CD46 receptor, which extends its cell tropism. Tyr481 can hydrogen bond to CD46 (Fig. 15.5e), an interaction responsible of the virus-receptor binding specificity. The use of an extended receptor-binding surface might explain MV recognition of multiple receptor molecules, which must be related to the efficient transmission of MV from host to host.

15.2.3 Carbohydrates as Viral Receptors

The cell surface displays a large variety of oligosaccharides linked to glycoproteins, proteoglycans and glycolipids, some of which are used by viruses to attach to host cells [61]. Most viral carbohydrate receptors are negatively charged and terminate the glycan moiety, features important for virus recognition. In some cases, however, viruses recognize neutral glycans such as histo-blood group antigens. Monomeric virus-carbohydrate interactions are usually of low affinity (mM range), however, virus particles have many recognition sites that easily engage several carbohydrate molecules, and thus attach to the cell surface with high avidity (Fig. 15.6a). Virus recognition of carbohydrates is associated also with functions other than cell attachment. Orthomyxo-, Paramyxo-, and Coronaviruses express envelope sialic acid-destroying glycoproteins (sialidases and esterases) that are essential for *in vivo* host infection. These enzymes can prevent re-attachment of newly released viruses from infected cells, can remove cell-bound viruses that fail to enter host cells, and can inhibit virus aggregation during budding. Sialidase inhibitors have proven to be useful anti-viral drugs for the treatment of some viral infections.

Sialic acid residues linked to glycoproteins and glycolipids act as receptors for many viruses. Sialic acids are derived from N-acetyl-neuraminic acid and mainly occupy the terminal position of a glycan chain, bound to a penultimate galactose through an (2–3) or (2–6) linkage [61], which renders them easily accessible. Moreover, sialic acids have a larger number of functional groups than other monosaccharides, and can thus participate in a network of polar and non-polar interactions with virus proteins (Fig. 15.6b) [62]. Sialic acids are receptors for distinct viruses, such as influenza, corona-, paramyxo-, toro-, adeno-, noro-, rota-, picorna-, parvo-, polyoma- and reovirus, some of which are important human pathogens [61]. Crystal structures of virus-sialic acid complexes showed that viruses use relatively recessed binding surfaces that interact mostly with the sialic acid face that bears a negatively charged carboxylate group (Figs. 15.6b, c, d) [62]. The mode of sialic acid recognition by distinct viruses can be similar, although viruses use structurally diverse proteins in receptor recognition. In some cases, however, related viruses such as murine polyomavirus and SV40 use similar capsid surface areas to recognize distinct motifs in the sialic acid molecules [5]. The virus-glycan contacts are thus not necessarily conserved (Fig. 15.6c, d).

Viral sialic acid receptors feature numerous cell-specific modifications that determine virus tropism, cell-to-cell transmission and pathogenicity. There are several examples of viruses responsible for serious diseases, such as adenovirus (Ad8, Ad19, Ad37) and enterovirus, which bind specifically to the (2–3)-linked sialic acid. The use of this sialic acid variant, which forms part of a branched glycan linked to the CD1a ganglioside, is responsible for the eye-tropism of those viruses, whose infection cause severe ocular diseases [29, 61]. Influenza A virus transmissibility and pathogenicity in humans also correlates with the recognition of specific sialic acid molecules by the envelope haemagglutinin (H) and neuraminidase (N)

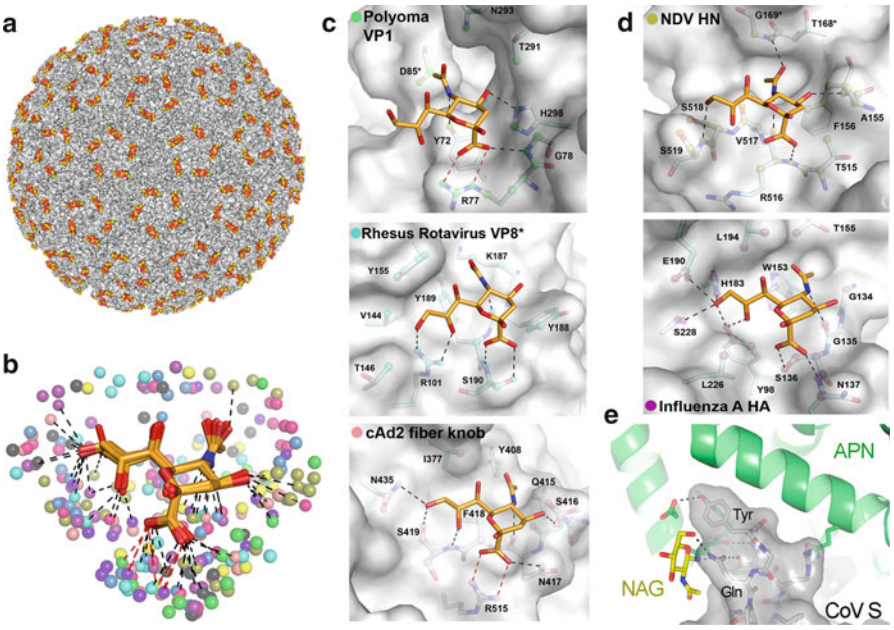


Fig. 15.6 Carbohydrates as viral receptors. (**a**) Multimeric binding of polyomavirus to its sialic acid receptor. Surface representation of the crystal structure (PDB ID 1SIE) of a polyomavirus particle (*surface in grey*) in complex with a sialic acid receptor fragment (*carbons in yellow*). (**b**) Possible contacts of viral proteins with terminal sialic acid, based on crystal structures [62]. Sialic acid is shown as sticks with carbons in orange. Viral atoms within 4.0 Å of the carbohydrate receptors in crystal structures of virus-sialic acid complexes are shown as spheres. Hydrogen bonds and salt bridges are indicated by *black* and *red lines*, respectively. (**c**) Non-enveloped viruses binding to sialic acid receptors. Detail of sialic acid binding interactions with polyoma VP1 (PDB ID 1VPS), rhesus rotavirus VP8* (PDB ID 1KQR) and cAd2 fibre knob (PDB ID 2WBV). (**d**) Enveloped viruses binding to sialic acid receptors. Network of interactions for the NDV-HN (PDB ID 1USR) and influenza HA protein (PDB ID 1HGG). (**e**) Coronavirus (CoV) binding to a glycan-containing region of its aminopeptidase N (APN) receptor. Detail of the interaction revealed by the crystal structure (PDB ID 4F5C) of a fragment of a porcine CoV spike protein (CoV S) in complex with the pig APN. CoV S is shown as *grey surface* and stick representations, the APN as *green ribbons*, with the key virus-binding N-acetyl-glucosamine (NAG) residue as sticks and *carbons in yellow*. The viral Tyr and Gln residues that contact the glycan in the CoV S protein are labelled. Hydrogen bonds are *dashed black lines*. Oxygens are *red* and nitrogens *dark blue* in all figures (Figures (**b-d**) were provided by U. Neu and T. Stehle, adapted with permission from [62])

proteins. The H (H1 to H15) and N (N1 to N9) protein variants identified in distinct species differ in their affinities for the (2–3)- or (2–6)-linked sialic acids, which is a key determinant in virus transmissibility [61]. Aquatic birds are the natural reservoir of influenza A, but the avian viruses are periodically transmitted to mammals, causing flu pandemics with significant numbers of deaths [63]. The H protein of avian influenza A (HA) binds preferentially to (2–3)-linked sialic acid receptors, which are not present in the human tracheal cells initially infected by influenza A [64];

transmission of avian influenza A from birds to humans or within the human population is therefore quite inefficient. Nonetheless, variants of some influenza A strains (H5N1) can infect humans, causing sporadic outbreaks with high mortality rates. Alterations in the virus receptor-binding specificity for (2–6)-linked sialic acids can cause the emergence of a pandemia such as the 1918 Spanish flu. In addition to these examples, other viruses can recognise specific substitutions in the sialic acid molecule [61].

Viruses can also bind to other negatively charged sulphated or neutral carbohydrates. Human herpesviruses are an example of viruses that attach to cell surface heparan sulphate (HS), and this interaction has been studied extensively for human herpes simplex viruses (HSV). HSV cell entry is complex and is mediated by several envelope glycoproteins (gC, gB, gD, gH and gL) [65]. The initial interaction of the virus particles with the cell is mediated by the gC and gB glycoproteins, which bind to HS. This interaction is nevertheless insufficient for virus entry, which also requires gD binding to cell surface proteins for virus-cell membrane fusion, catalysed by gB and the gH/gL complex. HS can thus be considered an HSV attachment factor that facilitates subsequent interactions with other receptors for cell entry.

Noroviruses exemplify a group of viruses that use neutral cell surface carbohydrate receptors, known as histo-blood group antigens [66]. Recognition is strain-specific and different virus strains recognize distinct types of antigens. These variations affect the tropism and pathogenesis of norovirus infections, as susceptibility to infection is dependent on the histo-blood group antigen receptors, which differ among individuals, who can therefore be resistant to infection by certain norovirus strains and susceptible to others. It is of interest to note that viruses can also recognize cell surface structures comprised by carbohydrates and amino acid residues, as recently described for some coronaviruses (CoV) [67]. A protruding loop with a tyrosine residue in the envelope spike of a subset of CoV docks between a neutral N-acetyl-glucosamine residue and an alpha helix exposed in the ectodomain of its receptor, the aminopeptidase N (APN) (Fig. 15.6e). This N-linked glycan is essential for virus binding to APN, and is a major determinant of the host range of some CoV [67].

Cell surface carbohydrates are ubiquitous, accessible and relatively variable "hooks" for cell attachment by highly distinct virus families. Carbohydrate-derived compounds can thus be used to inhibit many types of virus infections. Carbohydrates are major cell entry receptors for some viruses, and can be alternative or secondary entry receptors for others. Viruses that attach to protein receptors can evolve to use carbohydrates, thus expanding host cell tropism [4, 16, 68]. Enveloped viruses can also exploit the glycans linked to their envelope proteins to attach to cell surface lectins. Some of these lectins are specific for certain glycosylation patterns associated to virus membrane glycoproteins, as discussed above for DC-SIGN and HIV-1. Many enveloped viruses attach to host cells *via* DC-SIGN [6], a cell surface receptor for multiple pathogens. Glycans are thus intimately linked to virus-cell interaction.

15.3 Non-Enveloped Virus Entry into Host Cells: The Uncoating Process

15.3.1 Receptor-Mediated Uncoating of Picornaviruses

After binding to cell surface receptors, the viral genome must translocate from the particle to the cell cytoplasm (Fig. 15.1). In enveloped viruses, this transfer may be rather simple and occurs after fusion of the virus and cell membranes, as will be illustrated in Chap. 16. In non-enveloped viruses, mechanisms for genome penetration into the cytoplasm have not yet been characterized in detail, although several pieces of the process have been described. Particularly well-studied models include some viruses of the picornavirus family, PV and HRV, which will be discussed here.

The initial event during membrane penetration of picornaviruses is the opening of the viral capsid and the exit of the viral genome, a process known as uncoating. The capsids of PV and HRV are very similar and are characterized by a sedimentation coefficient of about 160S for PV and 150S for HRV; after RNA exit, the density of the empty capsids decreases markedly and their sedimentation coefficient is 80S [69–71]. Binding of soluble receptor molecules to the viral particles decreases their sedimentation coefficient to approximately 120S [72]. In addition, uncoating intermediates that lack the internal VP4 protein and have a sedimentation coefficient of about 135S have been described [69, 73, 74]. The distinct sedimentation behaviour of native virions and empty capsids allowed analysis of the uncoating process by ultracentrifugation in PV and HRV [70–72, 75, 76]. The structures of some of these capsid forms have been determined, and have helped to propose models for the structural rearrangements and dynamic processes associated with genome uncoating in these viruses (see below, Subsection 15.3.2).

The opening of the capsid in PV and HRV can be mediated by receptor binding, by low pH, or by the cooperative effect of both factors [71, 74–76]. Initial studies with PV and soluble poliovirus receptor (PVR) showed that receptor binding at physiological temperature (37 °C) mediates RNA exit from the capsid interior with no additional factors [70]. PVR binding to PV generates an intermediate particle with a sedimentation coefficient of 135S that lacks VP4, which is subsequently converted to empty 80S capsids. These particles are also observed during initial cell infection by PV [74], and thus represent entry intermediates. These findings show that PVR has an active role in PV entry; it is not just a “hook” used for attachment, but also an “unzipper” that mediates viral RNA uncoating [77]. Subsequent studies with HRV showed similar behaviour for the HRV receptor ICAM-1, which triggers RNA uncoating after binding to certain HRV serotypes at physiological temperatures [71, 72]. ICAM-1 binding to HRV serotypes 3 (HRV3) and 14 (HRV14) promotes RNA exit and formation of empty 80S capsids at neutral pH. This process is temperature dependent and requires temperatures over 25–30°C, necessary to overcome the high activation energy of the uncoating process (~45 kcal/mol) [72]; this energy is provided by receptor binding, an endothermic process in viruses sensitive to

receptor binding, such as PV and HRV3 [12]. PVR binding to PV is more endothermic than ICAM-1 binding to HRV3, which indicates that PVR is more efficient triggering uncoating than ICAM-1. As PV can enter host cells in the presence of agents that prevent endosomal acidification [74], RNA uncoating and transfer to the cytoplasm must therefore be mediated by receptor binding. PV infects hosts through the gastrointestinal tract; it is thus very stable at the low pH to which viruses are exposed during endocytosis. PV uses a receptor molecule suited to efficient uncoating at neutral pH and host cell entry.

The HRV are a group of viruses composed of around 100 different serotypes that bind to ICAM-1 (major group) or LDLR (minor group), as described above (Subsection 15.2.1). There is also certain diversity among serotypes in terms of the physiological factors that trigger HRV uncoating. Minor group HRV are sensitive to mildly acidic pH (5.5–6.0), which can trigger virus uncoating at physiological temperatures. Binding to LDLR does not mediate uncoating, and minor group HRV are thus dependent on endocytosis and endosome acidification for uncoating and host cell entry [16, 75]. The major group of HRV, which bind ICAM-1, comprises viruses that differ in stability and sensitivity to receptor binding [12, 71, 78]. Serotypes such as HRV3 are uncoated efficiently by ICAM-1 binding at neutral pH and physiological temperatures, whereas HRV16 are relatively stable and HRV14 have intermediate sensitivity. These differences in sensitivity to receptor binding are not related to affinity, but to binding energy. HRV3 binding to ICAM-1 is more endothermic than HRV16 binding [12]; the HRV3-receptor complex is thus less stable and is therefore primed for uncoating. On the contrary, HRV16-receptor complexes are more stable and require additional factors for uncoating, as explained below. In accordance with these findings, structural studies of HRV-receptor complexes prepared at physiological temperature show expansion of HRV3 following receptor binding and lack of expansion for HRV16 [79]. The energy absorbed by the HRV3-receptor interaction translates into capsid expansion, a metastable state primed for RNA exit.

The low pH at which cell surface receptor-bound viruses are exposed during endocytosis is a relevant factor that mediates virus uncoating (Fig. 15.1) [3]. Indeed, the mildly acidic pH (5.5–6.0) in endosomal compartments is optimal for uncoating of and infection by minor group HRV [16, 75, 80]. Exposure to low pH is also necessary for efficient uncoating and cell entry by some major group HRV [16, 76], although certain serotypes require only ICAM-1 binding, after which they can infect cells, even in the presence of agents that prevent endosome acidification. In the case of relatively stable HRV serotypes with pocket factor molecules, such as HRV16, both receptor binding and low endosome pH are needed for entry into host cells. The uncoating rate of HRV16 after binding ICAM-1 or exposure to mildly acidic pH (5.5–6.0) is very slow, and the virus can remain intact for several hours [71, 76]. After binding to ICAM-1, however, HRV16 particles become sensitive to low pH, which triggers rapid RNA release from the capsid [76]. Receptor binding appears to alter the HRV16 particles, priming them for low pH-mediated uncoating during endocytosis. This cooperative effect of the receptor and low pH in virus uncoating must be necessary to polarize RNA exit at the receptor-bound region of

the virus particle in endosomes, which must be relevant for efficient penetration of the cell membrane and infection.

The diverse receptor-mediated uncoating observed among major group rhinoviruses has been linked to the inherent stability and dynamics of the capsid. Biochemical studies of HRV indicate dynamic capsid behaviour [78, 81]; internal capsid polypeptides in the crystal structures, such as VP4 and the N terminus of VP1, are very sensitive to proteases, indicating that they are exposed to the environment at physiological temperatures. The dynamic behaviour of the capsids was termed "breathing" (see Chap. 6), and is likely to be mediated by the uncoordinated expansive-contractive movement of capsid protomers. Capsid breathing in HRV appears to be restricted by small molecules bound to a pocket in the VP1 [78, 81]. Natural molecules or pocket factors were identified as fatty acids in the capsids of the major group HRV16 and minor group HRV2. A group of molecules known as WIN compounds also bind to the VP1 pocket, restricting capsid dynamics and preventing uncoating. Moreover, the receptor-sensitive HRV3 serotype lacks pocket factor, which is present in the receptor-resistant HRV serotypes HRV16 and HRV2. These data suggest that capsid breathing is necessary for triggering receptor-mediated uncoating at physiological temperatures [79, 82].

15.3.2 The Structural Bases of Receptor-Mediated Virus Uncoating in Picornaviruses

Several structural studies over the last decade have examined the uncoating of PV and HRV intermediate particles by cryo-EM. One analysis determined the structure of the 135S PV particles, which lack VP4 but preserve the RNA inside, and compared this structure with 160S native virion and 80S empty capsids formed after RNA exit [73]. The 135S particle showed VP rearrangements relative to the native particle and 4 % expansion. Major rearrangements appear in the two-fold axes of the capsid and in the canyon region; however, the conformation of the five-fold axis, which was considered the RNA exit port, was similar to that of the 160S native PV capsid. HRV3 and ICAM-1 complexes prepared at 37 °C and in the process of uncoating were analysed by cryo-EM [79]; the structure also showed HRV3 capsid expansion not seen in stable HRV16-ICAM-1 complexes, indicating that capsid expansion must precede uncoating. The conformation of the RNA-containing HRV3 capsid in the virus-receptor complexes is similar to that of the native HRV capsid. Nevertheless, the empty HRV3 capsid is quite different, showing major rearrangements after RNA release; this has now been described in detail in a high-resolution structural study of empty HRV capsids [83]. Major differences between full and empty capsids in the HRV-receptor complexes appear at the five-fold axes; however, the structure of the five-fold axis regions are almost identical to those of native viruses in RNA-containing HRV3-receptor complexes [79]. These structural insights challenged the hypothesis of RNA exit through the

capsid five-fold axis, suggesting other exit ports. Recent studies with PV showed that RNA exit from the capsid near a two-fold axis, at a site that extends toward the receptor-binding canyon region [84].

The studies discussed here for PV and HRV show that receptors participate in the virus entry process, mediating not only virus-cell attachment but also RNA release, alone or with the contribution of low pH. The uncoating process requires multivalent receptor binding and capsid dynamics, which must explain the temperature dependence of this process. At physiological temperatures, capsids are dynamic entities whose protomers can move randomly, opening and closing holes at interprotomer junctions. This rapid, random movement does not allow for RNA release, probably because of the transient nature of the holes. After receptor binding at physiological temperatures, the capsid remains expanded and holes remain open for RNA release. PV and HRV recognize receptors *via* the canyon, a fragmented region at the junction of two protomers. PVR and ICAM-1 receptors can therefore act as wedges by binding at protomer junctions to maintain capsid expansion [12, 79], thus catalysing the uncoating process by locking the capsid in an intermediate, open state that allows RNA exit. The function of these two receptor molecules in virus uncoating is probably closely linked to the way they are recognized by PV and HRV.

The externalized genome in non-enveloped viruses either on the cell surface or in endosomes must translocate into the cell cytoplasm to initiate the intracellular phase of the life cycle. The study of the transfer of the genome from the virus particle to the cytoplasm is less understood in non-enveloped than in enveloped viruses. Structures of isolated envelope fusion proteins have delineated the process of membrane penetration or fusion of virus-cell membranes, described in Chap. 16. However, the analysis of viral genome penetration process in non-enveloped viruses requires the study of virus infecting host cells, which is methodologically challenging. Research carried out to date indicates two possible routes of penetration, either by membrane rupture or pores through which genome moves to the cytoplasm [16]. These disruptions in the membrane can be caused by hydrophobic capsid motif exposed during the uncoating process. The PV and HRV particles discussed here externalise the hydrophobic N-terminal region of VP1 that enable attachment of the particles to cell membranes in endosomes or on the cell surface. Moreover, the VP4 protein that is externalized together with the RNA bears a myristoyl group that can also interact with the cell membrane. The interaction of capsid motifs with the membrane can mediate its disruption for the transfer of the uncoated genome to the cytoplasm.

15.4 Perspectives and Conclusions

Viruses subvert cell surface molecules for cell entry and dissemination of infection. Receptor accessibility and cell expression patterns are important factors that underlie virus selection of specific surface receptors. Most of these receptors promote translocation of the viral genome into the cell. In some cases, however, viruses

attach to cellular factors that concentrate infectious particles on the plasma membrane and facilitate virus recognition of entry receptors. In other cases, viruses use cell surface molecules to spread the infection throughout the body; they can attach to migratory cells or induce signalling events that loosen cell-cell contacts and facilitate virus transmission. Cell surface molecules can thus have diverse functions in the dissemination of viral infections; the study of virus-receptor interactions is therefore of great relevance for understanding virus evolution, host tropism and pathogenesis.

The virus entry receptors have been the main focus of this chapter. These molecules are not just a "hook" for virus attachment, but can also participate in translocation of the viral genome into host cells, on the cell surface, or within endosomal compartments. Some virus receptors catalyse the entry process alone or in combination with other cell factors, such as the mildly acidic pH found in endosomes. Interaction with certain receptors transforms virus particles into metastable entities, priming delivery of their genomes to the cytoplasm. Alternatively, genome translocation can be mediated by low pH, but this requires endocytosis of receptor-bound viruses. The process of receptor-mediated virus entry continues to be studied. In some picornaviruses, discussed here, the energy absorbed in the virus-receptor interaction translates into virus particle expansion and genome exit. In other non-enveloped viruses, capsid expansion triggered by cell factors might be a necessary intermediate step for genome uncoating. Receptor binding also primes virus-cell membrane fusion in enveloped viruses, mediated by conformational changes in virus fusion proteins (Chap. 16). Cell surface molecules are thus key players in virus entry into host cells.

Virus-receptor interactions appear to be highly specific. There are many examples of virus selection of only a single member of closely-related cell surface molecules. This specificity is based on the recognition of certain unique structural features of a cell surface molecule, as well as on key polar interactions such as those described here for MV and its CD46 receptor. Virus-receptor surface complementarity must be important for virus recognition of certain receptor molecules. Viral proteins are shaped to dock into structural features of their receptors. Many viruses use concave surfaces for binding to cell surface receptors, which optimises receptor contact area and hides receptor-binding residues from antibodies. Nevertheless, viruses can also have protruding receptor-binding surfaces for receptor recognition. In both cases, viruses can escape from antibody neutralisation by mutating non-essential residues in these receptor-binding surfaces. Immune system pressure on viruses is a major determinant for the switch in receptor recognition observed in many virus groups, although this diversity can also be linked to opportunities to spread infection using ubiquitous or more accessible cell surface molecules. In addition to its high degree of specificity, virus-receptor recognition can be dynamic, and viruses can evolve to use alternative or distinct receptors, as illustrated here with several examples.

Preventing virus binding to receptors can efficiently block virus infection and cell damage. Receptor binding regions in virus proteins are relatively invariant and could potentially be targeted by immune responses to prevent infection. A large

number of structural studies have characterised viral antigenic sites targeted by neutralising antibodies. Many of these sites overlap receptor-binding regions, showing that blockade of virus binding to cell surface receptors is a major neutralisation mechanism. In viruses for which efficient vaccines have been developed (MV and PV), the receptor-binding residues are relatively well-exposed and accessible to antibodies. Viruses have nonetheless developed ways to evade neutralisation by hiding receptor-binding regions from antibodies; the use of concave and poorly accessible receptor-binding surfaces is one well-defined strategy of this type. In some viruses, these relatively inaccessible surfaces are also surrounded by variable loops and glycans, which further prevent antibody binding and neutralisation. Blocking infection of these viruses requires the design of improved vaccines that elicit immune responses focused on relatively inaccessible sites. These therapies require a deep understanding of virus-receptor interactions, including the determination of complex structures. These studies also open avenues for the development of molecules that prevent virus entry into cells. Soluble multimeric receptor molecules that impede virus binding to cells and infection were one of the first therapeutics developed for viruses such as HIV or HRV. Small molecule drugs have been developed, such as sialic acid analogues to treat influenza virus infection, or molecules that prevent HIV gp120 binding to its receptor, CD4 (see Chap. 20). The characterization of virus-receptor interactions is thus of considerable interest for the development of antiviral therapies.

Acknowledgements I thank U. Neu, T. Stehle, L. Xing and H.R. Cheng for providing figures included in this chapter and C. Mark for editorial assistance. Grant support by the National Institutes of Health (NIH P01-AI054456) and the Spanish Ministry of Science and Innovation (MICINN; BFU2011-23940) is acknowledged.

References and Further Reading

1. Marsh M, Helenius A (2006) Virus entry: open sesame. Cell 124:729–740
2. Grove J, Marsh M (2011) The cell biology of receptor-mediated virus entry. J Cell Biol 195:1071–1082
3. Mercer J, Schelhaas M, Helenius A (2010) Virus entry by endocytosis. Annu Rev Biochem 79:803–833
4. Baranowski E, Ruiz-Jarabo CM, Domingo E (2001) Evolution of cell recognition by viruses. Science 292:1102–1105
5. Stehle T, Casasnovas JM (2009) Specificity switching in virus-receptor complexes. Curr Opin Struct Biol 19:181–188
6. Backovic M, Rey FA (2012) Virus entry: old viruses, new receptors. Curr Opin Virol 2:4–13
7. Wang J-h (2002) Protein recognition by cell surface receptors: physiological receptors versus virus interactions. Trends Biochem Sci 27:122–126
8. Rossmann MG, He Y, Kuhn RJ (2002) Picornavirus-receptor interactions. Trends Microbiol 10:324–331
9. Mendelsohn CL, Wimmer E, Racaniello VR (1989) Cellular receptor for poliovirus: molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily. Cell 56:855–865

10. Greve JM, Davis G, Meyer AM, Forte CP, Yost SC, Marlor CW, Kamarck ME, McClelland A (1989) The major human rhinovirus receptor is ICAM-1. Cell 56:839–847
11. Staunton DE, Merluzzi VJ, Rothlein R, Barton R, Marlin SD, Springer TA (1989) A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. Cell 56:849–853
12. Xing L, Tjarnlund K, Lindqvist B, Kaplan G, Feigelstock D, Cheng RH, Casasnovas JM (2000) Distinct cellular receptor interactions in poliovirus and rhinoviruses. EMBO J 19:1207–1216
13. Xing L, Huhtala M, Pietiainen V, Kapyla J, Vuorinen K, Marjomaki V, Heino J, Johnson MS, Hyypia T, Cheng RH (2004) Structural and functional analysis of integrin alpha2I domain interaction with echovirus 1. J Biol Chem 279:11632–11638
14. Smith TJ, Chase ES, Schmidt TJ, Olson NH, Baker TS (1996) Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. Nature 383:350–354
15. Hewat EA, Verdaguer N, Fita I, Blakemore W, Brookes S, King A, Newman J, Domingo E, Mateu MG, Stuart DI (1997) Structure of the complex of an Fab fragment of a neutralizing antibody with foot-and-mouth disease virus: positioning of a highly mobile antigenic loop. EMBO J 16:1492–1500
16. Fuchs R, Blaas D (2010) Uncoating of human rhinoviruses. Reviews in Medical Virology 20:281–297
17. Verdaguer N, Fita I, Reithmayer M, Moser R, Blaas D (2004) X-ray structure of a minor group human rhinovirus bound to a fragment of its cellular receptor protein. Nat Struct Mol Biol 11:429–434
18. Bergelson JM, Chan M, Solomon KR, St John NF, Lin H, Finberg RW (1994) Decay-accelerating factor (CD55), a glycosylphosphatidylinositol-anchored complement regulatory protein, is a receptor for several echoviruses. Proc Natl Acad Sci USA 91:6245–6248
19. He Y, Lin F, Chipman PR, Bator CM, Baker TS, Shoham M, Kuhn RJ, Medof ME, Rossmann MG (2002) Structure of decay-accelerating factor bound to echovirus 7: a virus-receptor complex. Proc Natl Acad Sci USA 99:10325–10329
20. Pettigrew DM, Williams DT, Kerrigan D, Evans DJ, Lea SM, Bhella D (2006) Structural and functional insights into the interaction of echoviruses and decay-accelerating factor. J Biol Chem 281:5169–5177
21. Casasnovas JM, Markarian S, Hammar L (2004) Sensor surface interactions in the study of macromolecular assemblies. In: Cheng R, Hammar L (eds) Conformational proteomics of macromolecular architecture. World Scientific, Singapore
22. Zhang Y, Bergelson JM (2005) Adenovirus receptors. J Virol 79:12125–12131
23. Berk AJ (2007) Adenoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology, vol 2. Lippincott, Williams & Wilkins, Philadelphia, pp 2355–2394
24. Bewley MC, Springer K, Zhang Y-B, Freimuth P, Flanagan JM (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science 286:1579–1583
25. van Raaij MJ, Chouin E, van der Zandt H, Bergelson JM, Cusack S (2000) Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 A resolution. Structure 8:1147–1155
26. Walters RW, Freimuth P, Moninger TO, Ganske I, Zabner J, Welsh MJ (2002) Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. Cell 110:789–799
27. Persson BD, Reiter DM, Marttila M, Mei Y-F, Casasnovas JM, Arnberg N, Stehle T (2007) Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol 14:164–166
28. Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78:7727–7736
29. Nilsson EC, Storm RJ, Bauer J, Johansson SMC, Lookene A, Ångström J, Hedenström M, Eriksson TL, Frängsmyr L, Rinaldi S, Willison HJ, Domellöf FP, Stehle T, Arnberg N (2010)

The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17:105–109
30. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR (1993) Integrins αvβ3 and αvβ5 promote adenovirus internalization but not virus attachment. Cell 73:309–319
31. Wyatt R, Sodroski J (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science 280:1884–1888
32. Zhu P, Liu J, Bess J, Chertova E, Lifson JD, Grisé H, Ofek GA, Taylor KA, Roux KH (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature 441:847–852
33. Freed RO, Martin MA (2007) HIVs and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology, vol 2. Lippincott, Williams & Wilkins, Philadelphia, pp 2107–2185
34. Berger EA (1998) HIV entry and tropism. When one receptor is not enough. Adv Exp Med Biol 452:151–157
35. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648–659
36. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 393:705–711
37. Berger EA, Murphy PM, Farber JM (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17:657–700
38. Baribaud F, Pöhlmann S, Doms RW (2001) The role of DC-SIGN an DC-SIGNR in HIV and SIV attachment, infection, and transmission. Virology 286:1–6
39. Geijtenbeek TBH, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GCF, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR, Figdor CG, van Kooyk Y (2000) DC-SIGN, a dendritic cell-specific HIV-1 binding protein that enhances trans-infection of T cells. Cell 100:587–597
40. Feinberg H, Mitchell DA, Drickamer K, Weis WI (2001) Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. Science 294:2163–2166
41. Guo Y, Feinberg H, Conroy E, Mitchell DA, Alvarez R, Blixt O, Taylor ME, Weis WI, Drickamer K (2004) Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. Nat Struct Mol Biol 11:591–598
42. Puryear WB, Yu X, Ramirez NP, Reinhard BM, Gummuluru S (2012) HIV-1 incorporation of host-cell-derived glycosphingolipid GM3 allows for capture by mature dendritic cells. Proc Natl Acad Sci USA 109:7475–7480
43. Izquierdo-Useros N, Lorizate M, Contreras FX, Rodriguez-Plata MT, Glass B, Erkizia I, Prado JG, Casas J, Fabrias G, Kräusslich H-G, Martinez-Picado J (2012) Sialyllactose in viral membrane gangliosides is a novel molecular recognition pattern for mature dendritic cell capture of HIV-1. PLoS Biol 10:e1001315
44. Lamb RA, Parks GD (2007) Paramyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds) Fields virology, vol 1. Lippincott, Williams & Wilkins, Philadelphia, pp 1449–1496
45. Lamb RA, Paterson RG, Jardetzky TS (2006) Paramyxovirus membrane fusion: lessons from the F and HN atomic structures. Virology 344:30–37
46. Plemper RK, Brindley MA, Iorio RM (2011) Structural and mechanistic studies of measles virus illuminate paramyxovirus entry. PLoS Pathog 7:e1002058
47. Bose S, Zokarkar A, Welch BD, Leser GP, Jardetzky TS, Lamb RA (2012) Fusion activation by a headless parainfluenza virus 5 hemagglutinin-neuraminidase stalk suggests a modular mechanism for triggering. Proc Natl Acad Sci USA 109:E2625–E2634
48. Negrete OA, Levroney EL, Aguilar HC, Bertolotti-Ciarlet A, Nazarian R, Tajyar S, Lee B (2005) EphrinB2 Is the entry receptor for nipah virus, an emergent deadly paramyxovirus. Nature 436:401–405

49. Bonaparte MI, Dimitrov AS, Bossart KN, Crameri G, Mungall BA, Bishop KA, Choudhry V, Dimitrov DS, Wang LF, Eaton BT, Broder CC (2005) Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. Proc Natl Acad Sci USA 102:10652–10657
50. Negrete OA, Wolf MC, Aguilar HC, Enterlein S, Wang W, Muhlberger E, Su SV, Bertolotti-Ciarlet A, Flick R, Lee B (2006) Two key residues in ephrinB3 are critical for its use as an alternative receptor for Nipah virus. PLoS Pathog 2:e7
51. Dorig RE, Marcil A, Chopra A, Richardson CD (1993) The human CD46 molecule is a receptor for measles virus (Edmonston strain). Cell 75:295–305
52. Naniche D, Varior-Krishman G, Cervoni F, Wild TF, Rossi B, Rabourdin-Combe C, Gerlier D (1993) Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. J Virol 67:6025–6032
53. Tatsuo H, Ono N, Tanaka K, Yanagi Y (2000) SLAM (CDw150) is a cellular receptor for measles virus. Nature 406:893–897
54. Muhlebach MD, Mateo M, Sinn PL, Prufer S, Uhlig KM, Leonard VHJ, Navaratnarajah CK, Frenzke M, Wong XX, Sawatsky B, Ramachandran S, McCray PB, Cichutek K, von Messling V, Lopez M, Cattaneo R (2011) Adherens junction protein nectin-4 is the epithelial receptor for measles virus. Nature 480:530–533
55. Noyce RS, Bondre DG, Ha MN, Lin L-T, Sisson G, Tsao M-S, Richardson CD (2011) Tumor cell marker PVRL4 (nectin 4) is an epithelial cell receptor for measles virus. PLoS Pathog 7: e1002240
56. Crennell S, Takimoto T, Portner A, Taylor G (2000) Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. Nat Struct Biol 7:1068–1074
57. Bowden TA, Aricescu AR, Gilbert RJ, Grimes JM, Jones EY, Stuart DI (2008) Structural basis of nipah and hendra virus attachment to their cell-surface receptor ephrin-B2. Nat Struct Mol Biol 15:567–572
58. Xu K, Rajashankar KR, Chan YP, Himanen JP, Broder CC, Nikolov DB (2008) Host cell recognition by the henipaviruses: crystal structures of the Nipah G attachment glycoprotein and its complex with ephrin-B3. Proc Natl Acad Sci U S A 105:9953–9958
59. Santiago C, Celma ML, Stehle T, Casasnovas JM (2010) Structure of the measles virus hemagglutinin bound to the CD46 receptor. Nat Struct Mol Biol 17:124–129
60. Hashiguchi T, Ose T, Kubota M, Maita N, Kamishikiryo J, Maenaka K, Yanagi Y (2011) Structure of the measles virus hemagglutinin bound to its cellular receptor SLAM. Nat Struct Mol Biol 18:135–141
61. Olofsson S, Bergström T (2005) Glycoconjugate glycans as viral receptors. Ann Medicine 37:154–172
62. Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: rules of engagement. Curr Opin Struct Biol 21:610–618
63. Lipatov AS, Govorkova EA, Webby RJ, Ozaki H, Peiris M, Guan Y, Poon L, Webster RG (2004) Influenza: emergence and control. J Virol 78:8951–8959
64. Matrosovich MN, Matrosovich TY, Gray T, Roberts NA, Klenk H-D (2004) Human and avian influenza viruses target different cell types in cultures of human airway epithelium. Proc Natl Acad Sci USA 101:4620–4624
65. Navaratnarajah CK, Miest TS, Carfi A, Cattaneo R (2012) Targeted entry of enveloped viruses: measles and herpes simplex virus I. Curr Opin Virol 2:43–49
66. Huang P, Farkas T, Marionneau S, Zhong W, Ruvoen-Clouet N, Morrow AL, Altaye M, Pickering LK, Newburg DS, LePendu J, Jiang X (2003) Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. J Infect Dis 188:19–31
67. Reguera J, Santiago C, Mudgal G, Ordoño D, Enjuanes L, Casasnovas JM (2012) Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLOS Pathog 8:e1002859
68. Li F (2012) Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J Virol 86:2856–2858

69. Rueckert RR (1996) Picornaviridae: the viruses and their replication. In: Fields BN, Knipe DM, Howley PM, Melnick JL, Chanock RM, Roizman B, Monath TP (eds) Virology, vol 1. Raven, New York, pp 609–654
70. Kaplan G, Freistadt MS, Racaniello VR (1990) Neutralization of poliovirus by cell receptors expressed in insect cells. J Virol 64:4697–4702
71. Hoover-Litty H, Greve JM (1993) Formation of rhinovirus-soluble ICAM-1 complexes and conformational changes in the virion. J Virol 67:390–397
72. Casasnovas JM, Springer TA (1994) The pathway of rhinovirus disruption by soluble intercellular adhesion molecule 1 (ICAM-1): An intermediate in which ICAM-1 is bound and RNA is released. J Virol 68:5882–5889
73. Belnap DM, Filman DJ, Trus BL, Cheng N, Booy FP, Conway JF, Curry S, Hiremath CN, Tsang SK, Steven AC, Hogle JM (2000) Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus. J Virol 74:1342–1354
74. Hogle JM (2002) Poliovirus cell entry: common structural themes in viral cell entry pathways. Ann Rev Microbiol 56:677–702
75. Prchla E, Kuechler E, Blaas D, Fuchs R (1994) Uncoating of human rhinovirus serotype 2 from late endosomes. J Virol 68:3713–3723
76. Nurani G, Lindqvist B, Casasnovas JM (2003) Receptor priming of major group human rhinoviruses for uncoating and entry at mild low-pH environments. J Virol 77:11985–11991
77. Racaniello VR (1996) The poliovirus receptor: a hook, or an unzipper? Structure 4:769–773
78. Lewis JK, Bothner B, Smith TJ, Siuzdak G (1998) Antiviral agent blocks breathing of the common cold virus. Proc Natl Acad Sci USA 95:6774–6778
79. Xing L, Casasnovas JM, Cheng RH (2003) Structural analysis of human rhinovirus complexed with ICAM-1 reveals the dynamics of receptor-mediated virus uncoating. J Virol 77:6101–6107
80. Schober D, Kronenberger P, Prchla E, Blaas D, Fuchs R (1998) Major and minor receptor group human rhinoviruses penetrate from endosomes by different mechanisms. J Virol 72:1354–1364
81. Roy A, Post CB (2012) Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity. Proc Natl Acad Sci USA 109:5271–5276
82. Casasnovas JM (2000) The dynamics of receptor recognition by human rhinoviruses. Trends Microbiol 8:251–254
83. Garriga D, Pickl-Herk A, Luque D, Wruss J, Caston JR, Blaas D, Verdaguer N (2012) Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. PLoS Pathog 8:e1002473
84. Bostina M, Levy H, Filman DJ, Hogle JM (2011) Poliovirus RNA is released from the capsid near a twofold symmetry axis. J Virol 85:776–783

Further Reading

Especially recommended for further reading are the following references listed above: References [1–3] present a general overview of virus-receptor recognition and virus entry into host cells; references [4, 5] describe virus switch of receptor specificity; reference [6] is part of a *Current Opinion in Virology* issue describing recent research on virus entry. I also recommend the following references related to specific chapter sections: 15.2.1, references [8, 16] (picornavirus), and [22] (adenovirus); 15.2.2, references [34, 35] (HIV-1), and [45, 46] (paramyxovirus); 15.2.3, references [61, 62]; 15.3, references [16, 74, 79].

Chapter 16
Entry of Enveloped Viruses into Host Cells: Membrane Fusion

Vicente Más and José A. Melero

Abstract Viruses are intracellular parasites that hijack the cellular machinery for their own replication. Therefore, an obligatory step in the virus life cycle is the delivery of the viral genome inside the cell. Enveloped viruses (*i.e.*, viruses with a lipid envelope) use a two-step procedure to release their genetic material into the cell: (i) they first bind to specific surface receptors of the target cell membrane and then, (ii) they fuse the viral and cell membranes. This last step may occur at the cell surface or after internalization of the virus particle by endocytosis or by some other route (*e.g.*, macropinocytosis). Remarkably, the virus-cell membrane fusion process goes essentially along the same intermediate steps as other membrane fusions that occur for instance in vesicular fusion at the nerve synapsis or cell-cell fusion in yeast mating. Specialized viral proteins, fusogens, promote virus-cell membrane fusion. The viral fusogens experience drastic structural rearrangements during fusion, liberating the energy required to overcome the repulsive forces that prevent spontaneous fusion of the two membranes. This chapter describes the different types of viral fusogens and their mode of action, as are currently known.

Keywords Class I fusion protein • Class II fusion protein • Class III fusion protein • Enveloped virus • Fusion pore • Glycoprotein • Membrane • Membrane fusion intermediate • Post-entry events • Viral fusogen • Virus entry

V. Más • J.A. Melero (✉)
Biología Viral, Centro Nacional de Microbiología and CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Majadahonda, 28220 Madrid, Spain
e-mail: jmelero@isciii.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_16,

Abbreviations

FP	Fusion peptide
G	Glycoprotein
GPI	Glycosyl phosphotidylinositol
HA	Haemagglutinin
HIV	Human immunodeficiency virus
hMPV	Human metapneumovirus
HN	Haemagglutin-neuraminidase
HRA and HRB	Heptad repeat sequences A and B, respectively
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
RSV	Respiratory syncytial virus
SFV	Semliki forest virus
SNARE	Soluble N-ethylamine sensitive factor attachment receptor protein
TBV	Tick-borne encephalitis virus
TM	Transmembrane
VSV	Vesicular stomatitis virus

16.1 Introduction

Enveloped viruses are characterized by having a lipid bilayer (envelope) surrounding the virus particle (virion) (see Chap. 11). One or several virus-encoded glycoproteins are inserted into the envelope. These proteins are exposed at the virion surface and are responsible therefore of the initial interactions of the virus with the target cell, leading to "virus entry". In fact, it is only the cargo inside the envelope layer and not the virus itself what is actually discharged inside the cell. The virus glycoproteins are also the main targets of the neutralizing antibody response produced by the host in its defense against the virus.

To infect a new cell, the virus particle must first attach to the cell surface through non-covalent interactions of one or more of the viral glycoproteins anchored into the lipid bilayer with specific cell surface receptors. These interactions are described in detail in Chap. 15. Suffice to say here that virus-receptor binding is one of the factors that can influence virus tropism; *i.e.*, which cell types are actually infected by the virus.

For enveloped viruses, fusion of the viral and cell membranes is an obligatory step that follows virus binding to cells. Virus-cell fusion is therefore the step at which the virus particle loses its individuality. Membrane fusion may proceed at the cell surface or alternatively after internalization of the virus particle, generally by endocytosis. In either case, fusion is driven by specialized viral glycoproteins (fusogens) which are activated (triggered) by specific events occurring either at the cell surface or inside the endosome. The viral fusogens are in metastable

conformations in the virus particle. Once triggered, they initiate a series of conformational changes (in most cases irreversibly) that facilitate approximation of the two membranes, followed by fusion. At the end of the fusion process, the viral fusogens adopt highly stable conformations. The free energy liberated during the transition from the metastable pre-fusion to the highly stable post-fusion conformation drives the fusion process.

16.2 General Principles of Membrane Fusion

16.2.1 Protein-Free Membrane Fusion

Lipid mixing occurs spontaneously in monolayers but several forces prevent the spontaneous mixing of lipids between bilayer membranes [1]. Most important among them are: (i) hydrophobic effects that seek to minimize solvent-exposed apolar surfaces, (ii) elastic forces that prevent monolayer deformation and (iii) electrostatic repulsions between negatively charged phospholipids. Nevertheless, fusion between protein-free lipid bilayers (*e.g.*, liposomes) can be induced under certain conditions. For instance, certain phospholipids (*e.g.*, phosphatidylcholine, PC) induce positive curvature of the lipid monolayer whereas others (*e.g.*, phosphatidylethanolamine, PE) induce negative curvature. The distribution of PC and PE between the two leaflets of the lipid bilayer can either promote or inhibit spontaneous protein-free membrane fusion. Also direct dehydration between bilayers promotes fusion by bringing the two membranes into very close contact. However, under most physiological conditions, specialized proteins are needed to overcome the repulsive forces that prevent membrane fusion.

Independently of the driving machinery, fusion of two lipid bilayers occurs in a stepwise manner that includes the formation of an hourglass-like structure, known as the lipid stalk (Fig. 16.1a) [2]. This stalk is then expanded forming what is called the hemifusion diaphragm, in which lipids of the two distal leaflets of the bilayers are now in direct contact. Finally, rupture of the hemifusion diaphragm leads to formation of the fusion pore that is then expanded to complete membrane fusion and content mixing of the two compartments. However, opening of the fusion pore may be a reversible step and does not always lead to full fusion [3]. Energy is therefore required through all steps of the fusion process, including expansion of the fusion pore. Hence, proteins when present must operate from the initial stages of membrane deformation to final merging of the two membranes.

16.2.2 Virus-Induced Membrane Fusion

Enveloped viruses contain specialized surface glycoproteins that mediate: (i) initial binding of virus to the cell surface and (ii) fusion of the virus and cell membranes,

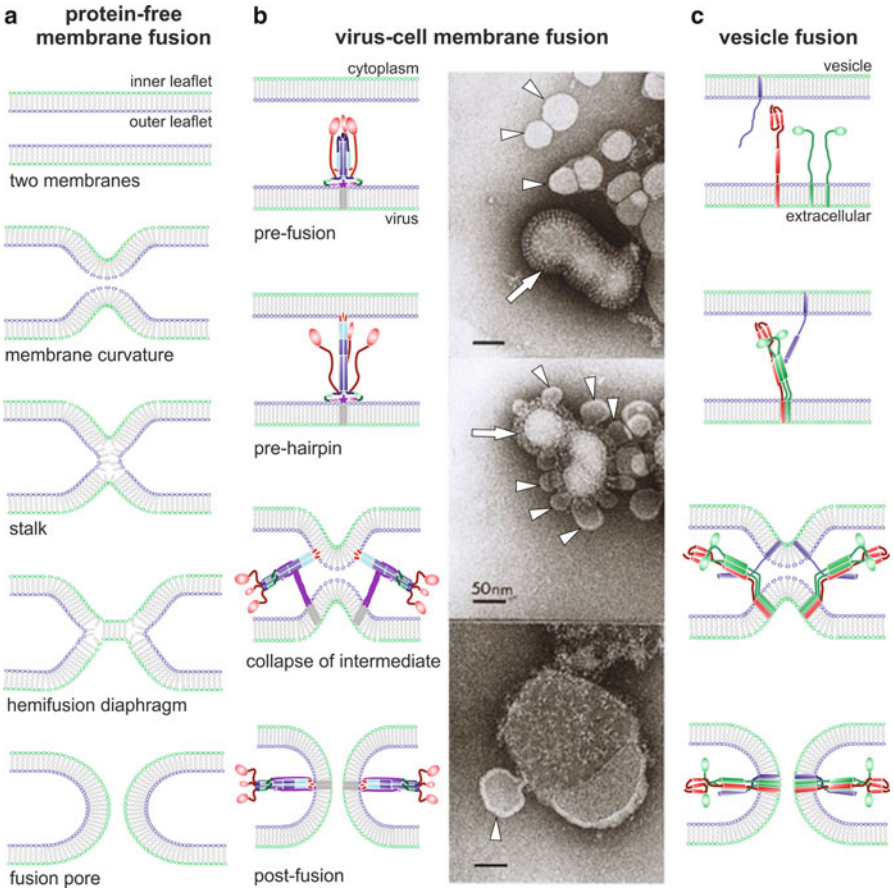


Fig. 16.1 Steps of the membrane fusion process. **(a)** Diagram of the fusion steps between two protein-free lipid bilayers. From *top* to *bottom*: the lipids (represented by heads and tails) of the two bilayers are initially curved into nipple-like structures that approach the two membranes. This is followed by formation of the stalk in which the two proximal leaflets are fused. This stalk is then expanded forming a hemifusion diaphragm in which lipids of the distal leaflets of the bilayers are now in direct contact. Finally, rupture of the hemifusion diaphragm leads to formation of the fusion pore. **(b)** Diagram of the virus-cell fusion process: As an example, fusion mediated by the influenza HA is illustrated. *Left panels:* HA is a homotrimer that initially binds to the cell surface (not shown) by interactions of each subunit head with sialic acid. Then, the virus is internalized by endocytosis. For simplicity, only 1 HA trimer in the pre-fusion conformation is shown, anchored into the viral membrane. After endosome acidification, the HA globular head falls apart, allowing refolding of the molecule to produce three long α-helices. The fusion peptides, placed at the N-terminal end of each α-helix, insert into the target membrane. This intermediate, dubbed pre-hairpin, refolds to bring the two membranes into proximity leading to formation of the lipid stalk followed by formation of the hemifusion diaphragm (not shown). Finally, the fusion pore is formed by the concerted action of several HA molecules that adopt a very stable post-fusion conformation. *Right panels:* The upper panel shows a mixture of an influenza virus particle (strain X31, H3N2, *white arrow*) and liposomes (some of them indicated by *white arrowheads*) made with lipids commonly found in cell membranes, incubated at neutral pH. Note the glycoprotein spikes

either at the cell surface or after endocytosis [4]. The initial attachment of the virus to the cell surface may involve only one type of viral glycoprotein (*e.g.*, influenza virus) or it may require the combined action of several viral glycoproteins (*e.g.*, herpesvirus). However, the actual process of membrane fusion is driven by a specialized type of viral glycoproteins (the viral fusogens) which may or may not have participated additionally in attachment.

Although the sequences of viral fusion proteins vary considerably, they all share certain structural characteristics and are subject to analogous structural rearrangements during the fusion process. As an example, Fig. 16.1b (left panels) illustrates the structural changes that the influenza virus haemagglutin (HA) undergo during fusion of the viral and the endosomal membrane. In this case, influenza HA also mediates the initial interaction of the virus with sialic acid of glycoproteins or glycolipids at the cell surface (see Chap. 15). After this initial binding, the virus is internalized by endocytosis. Acidification of the endosome (probably by fusion with lysosomes) triggers HA to start the conformational changes depicted in Fig. 16.1b (left panels) and described next.

Independently of the triggering event, all viral fusion proteins undergo conformational changes upon activation that lead to the formation of an extended unstable intermediate, dubbed pre-hairpin (Fig. 16.1b, "pre-hairpin"). Formation of the pre-hairpin intermediate involves very large-scale structural rearrangements in the viral fusogens with exposure of hydrophobic segments or loops (the fusion peptide, FP). Since the hydrophobic fusion peptide cannot be exposed to a hydrophilic environment it inserts into the target membrane. At this point the viral and target membranes are bridged by two separate segments of the same polypeptide; one is the fusion peptide bound to the target membrane and the other is the transmembrane (TM) region of the viral fusogen inserted into the viral membrane. The pre-hairpin intermediate may have a relatively long half-life; for the human immunodeficiency

Fig. 16.1 (continued) (mostly haemagglutinin, HA) sticking out of the viral membrane in contrast with the smooth surface of liposomes. The middle panel shows the same virus/liposome mixture after incubation for 5–10 s at pH 5.0 followed by neutralization. Note binding of liposomes to the virus surface and initiation of virus-liposome fusion. The lower panel shows the virus/liposome mixture after incubation for 5 min at pH 5.0 followed by neutralization. Note that the virus has fused with several liposomes, yielding a large vesicle with viral glycoproteins disperse throughout the surface and with a small liposome still in the process of fusion. The HA spikes also have changed morphology after exposure to low pH and fusion. (Courtesy of L.J. Calder and S.A. Wharton, Division of Virology, MRC National Institute for Medical Research, London, UK). (**c**) Diagram of vesicle fusion at the synaptic junction: Initially one of the SNARE proteins (synaptobrevin, *blue*) is inserted into the vesicle membrane, while three other SNAREs (two SNAP 25, *green* and one syntaxin, *red*) are inserted in the plasma membrane. After an initial interaction, refolding of the SNAREs leads to formation of a bundle of four parallel α-helices that drives approximation of the two membranes and formation of the stalk and hemifusion intermediates (not shown). Completion of the SNARE complex results in formation of the fusion pore. In the two lower panels of parts (b) and (c), two HAs and two SNARE complexes are shown surrounding the fusion pore, although the actual number of molecules involved in fusion pore formation is likely to be higher

virus (HIV) gp41 protein, the half-life seems to be several minutes [5], but in other cases, it may only be a few seconds [6]. The pre-hairpin bridge then collapses bringing into close apposition the viral and target membranes, which are distorted probably into protein-free nipple-like configurations (Fig. 16.1b, "collapse of intermediate"). This is followed by the formation of a lipid stalk and a hemifusion diaphragm (in analogy with the protein-free membrane fusion path), which allows lipid mixing between the two proximal leaflets of the viral and target membranes. Finally, the hemifusion diaphragm opens to form a transient fusion pore that may flicker open and closed until it expands [7], leaving the viral fusogen in a highly stable post-fusion hairpin conformation inserted into the target membrane (Fig. 16.1b, "post-fusion").

Influenza HA dependent membrane fusion can be reproduced in the test tube with purified virus and liposomes and observed by electron microscopy, as illustrated in Fig. 16.1b (right panels). After mixing of virions (arrows) and liposomes (arrowheads) they remain separated (upper panel), since the former lack the influenza virus receptor (sialic acid). However, a brief pulse at low pH (middle panel) exposes the fusion peptide of the influenza HA which is then inserted into the membrane of multiple liposomes. Longer pulses of low pH result in fusion of the virus with multiple liposomes leading to formation of large vesicles (lower panel; note two vesicles, one small and one large, caught in the process of fusion).

The fusion peptides probably insert only into the outer leaflet of the cell target membrane. Due to the large number of fusogen molecules present at the viral surface, multiple fusion peptides may interact with the external leaflet of the target membrane upon formation of the pre-hairpin intermediate, potentially initiating membrane deformation. This suggests that cooperativity between several viral fusogens may be required for membrane fusion. In fact, fusion mediated by the influenza HA is positively affected by protein density [8]. It is estimated that 4–6 HA molecules are required for fusion, forming a protein ring at the periphery of the fusion pore. Also, electron microscopy (Chap. 3) and X-ray crystallography (Chap. 4) results indicate that the E1 glycoprotein of Semliki Forest alphavirus interacts cooperatively during membrane insertion and fusion [9].

Despite the above arguments in favor of cooperativity, calculations of the energy barrier that must be overcome en route to a hemifusion diaphragm is estimated to be about 40–50 kcal.mol^{-1}. A free energy of roughly this magnitude could be recovered from the collapse of one or two pre-hairpin intermediates, depending on the interactions driving such collapse. In fact, experiments with HIV suggest that only one or two active envelope glycoproteins are sufficient for fusion [10], although later estimates have increased this number [11]. It may be that the fusion proteins of HIV and other retroviruses have evolved to manage with a single fusion protein, as the number of envelope glycoproteins in the virus particle (estimated 15–20, in contrast to hundreds in other viruses) is rather sparse.

Formation of the pre-hairpin structure and refolding of this intermediate entails some of the most drastic protein rearrangements ever found in biology. Pre-hairpin collapse involves folding back of the membrane proximal domain of the viral

fusogen onto a trimeric core whose distal end from the viral membrane is inserted into the target membrane (Fig. 16.1b). Zippering together of these two domains brings the membranes into close proximity. Dehydration of the initial contact site induces monolayer rupture resulting in lipid stalk formation and hemifusion. However, formation of the fusion pore requires further structural rearrangements, including interactions between regions adjacent to the fusion peptide and the transmembrane region [12, 13] and, probably, additional contacts between these two hydrophobic regions that are now inserted into the same membrane. For instance, membrane fusion by the influenza HA with a glycosyl phosphotidylinositol (GPI) anchor replacing the TM region halts at the hemifusion stage [14].

Finally, enlargement of the initial fusion pore is probably the most energy demanding step and requires the coordinated action of several fusogen molecules that surround the early nipple-like fusion intermediate [15].

16.2.3 Vesicle and Cell-Cell Fusion

This topic is brought here only to emphasize the analogies and differences between membrane fusions promoted by unrelated proteins. Vesicle fusion is required for essential biological processes, such as exocytosis and synaptic transmission. Cell-cell fusion is involved in hypodermal cell fusion in *C. elegans*, sperm-egg fusion, yeast mating (mating of two haploid yeast cells to produce a diploid cell), placenta formation in mammals, and muscle and bone formation.

In all cases, membrane fusion follows the same steps already described in previous sections; *i.e.*, deformation and approximation of the two membranes, formation of the stalk and hemifusion intermediates and finally formation and enlargement of the fusion pore (Fig. 16.1c). In analogy with virus-cell fusion, vesicle and cell-cell fusion requires formation of highly stable protein assemblies that provide the energy necessary to overcome the repulsive forces of membranes in close proximity [16]. Also, vesicle and cell-cell fusion, as viral fusion, requires higher order multimerization of the fusogens that delineate the hemifusion diaphragms and the fusion pores [17].

The main difference between virus-cell fusion and vesicle or cell-cell fusion is that in the former process the protein fusogen is present only in the viral membrane. In contrast, the proteins involved in vesicle fusion and cell-cell fusion are initially inserted in the two membranes predestined to fuse (Fig. 16.1c). In synaptic vesicles, the main proteins responsible of membrane fusion are the so-called SNARE (soluble N-ethylamine sensitive factor attachment receptor protein) proteins [18] which share a conserved 60–70 amino acid motif. These proteins, when they find each other refold into a highly stable four-helix parallel coiled-coil bundle that resembles the six-helix bundle formed by the heptad repeats (structural motifs with a repeating pattern of seven amino acids) of certain viral fusogens (see below). Formation of the four-helix bundle leads to membrane apposition and hemifusion, as with the collapse of the pre-hairpin intermediate of viral fusogens. However, a

unique characteristic of vesicle fusion is that the protein machinery involved in the process is disassembled, once fusion is finished to be reused in subsequent fusion events. This is accomplished by the ATPase (adenosine triphosphatase) N-ethylmaleimide sensitive factor (NSF) [19].

In contrast to vesicle fusion, cell-cell fusion entails the same set of fusion proteins in the two membranes. For instance, the exceptional process of hypodermal cell fusion in *C. elegans* to form a large multinucleated syncytium of all skin cells is driven by the EFF-1 protein [20]. Unlike SNAREs and viral fusogens, EFF-1 has a homotypic fusion machinery in the opposite membrane. In other words, both membranes must have EFF-1 for fusion to occur. Nevertheless, cell-cell fusion is a multistep process that goes along the same lipid intermediates as viral and vesicle fusion.

16.3 Viral Fusion Proteins

Based on biosynthetic and structural characteristics, viral fusogens have been classified into three categories (Table 16.1). Class I fusion glycoproteins are characterized by being synthesized as inactive precursors that require proteolytic processing to become fusion-competent. They are all homotrimers that upon fusion refold into hairpins containing a long central coiled-coil core structure (formed by helices that are coiled together). Class II fusion glycoproteins are derived from longer polyprotein precursors that are proteolytically processed during biosynthesis. The class II fusion proteins form icosahedral scaffolds of protein dimers at the viral surface. During fusion, these proteins undergo an oligomeric rearrangement, converting the metastable prefusion dimer into a stable hairpin homotrimer composed of β-sheet structures. Finally, class III glycoproteins are not proteolytically processed. Their post-fusion hairpin trimer displays a central α-helical coiled-coil, as class I glycoproteins, but the fusion domain exposes two fusion loops located at the tip of an elongated β-sheet, revealing a striking convergence with class II fusion proteins.

16.3.1 Class I Viral Fusion Proteins

The first atomic structure of any viral or cellular glycoprotein was determined by X-ray crystallography and reported in 1981 by the laboratories of Wiley and Skehel [21]. It was the structure of the influenza haemagglutinin (HA) trimeric ectodomain (the domain that protrudes from the plasma membrane), as released from the virus particles by bromelain treatment, which cleaves the HA polypeptides near the TM region.

Influenza HA is synthesized in the infected cell as a polypeptide precursor (HA0) of about 550 amino acids that is cleaved proteolytically to generate the HA1 (roughly the N-terminal two thirds) and HA2 (the C-terminal third) chains that

Table 16.1 Classification of viral fusion proteins

Class	Virus family	Representative	Viral fusogen	Involved in attachment
I	*Orthomyxoviridae*	Influenza virus	Haemaglutinin (HA)	Yes
	Retroviridae	Human immunodeficiency virus (HIV)	Envelope glycoprotein; gp 41 subunit	Yes
	Filoviridae	Ebola virus	GP glycoprotein	Yes
	Coronaviridae	Severe acute respiratory syndrome (SARS) virus	S glycoprotein	Yes
	Paramyxoviridae	Sendai virus	F glycoprotein	No
II	*Alphaviridae*	Semliki Forest Virus	E1 glycoprotein	No
	Flaviviridae	Dengue virus	E glycoprotein	Yes
III	*Rhabdoviridae*	Vesicular stomatitis virus	G glycoprotein	Yes
	Baculoviridae	Baculovirus	Gp64 glycoprotein	Yes
	Herpesviridae	Herpes simplex virus	gB glycoprotein	No

remain covalently linked by a disulfide bond. At the newly created HA2 N-terminus there is a stretch of hydrophobic amino acids, called the fusion peptide, which is inserted into the target membrane during fusion. The overall structure of the influenza HA is that of an elongated spike sticking out of the membrane. The distal head, formed exclusively by HA1 sequences, bears the receptor (sialic acid) binding site, formed by a shallow pocket exposed on its outward-forming surface. The stem, made largely by HA2 amino acids, is a trimeric α-helical coiled-coil. The structural rearrangements of the influenza HA during membrane fusion are shown in Fig. 16.1b.

As influenza, other viruses also contain class I fusion glycoproteins that have both receptor and membrane fusion activities (Table 16.1). For instance, the envelope glycoprotein of HIV that is also proteolitically processed and that binds to protein receptors (CD4) and chemokine co-receptors before engaging in membrane fusion at the cell surface. Similarly, the receptor-binding proteins of filovirus and coronavirus mediate additionally viral-cell membrane fusion.

In contrast, the attachment and fusion activities reside in two different surface glycoproteins of paramyxoviruses. The attachment protein (named HN, H or G) is required for the initial interaction of the virus with the cell surface (see Chap. 15 for virus receptor usage). Once the virus is bound to the cell, the other major viral glycoprotein (called F, for fusion) is triggered to promote fusion of the viral and cell membranes. Structure determination of prototypic paramyxovirus F proteins in the pre-fusion metastable conformation [22] and in the post-fusion state [23] by X-ray crystallography, as well as identification of fusion intermediates [24], has provided the most complete picture of the membrane fusion process driven by class I fusion glycoproteins, as depicted in Fig. 16.2.

The paramyxovirus F protein, as other class I glycoproteins, is synthesized as an inactive precursor (F0) that is translocated co-translationally to the lumen of the endoplasmic reticulum where it assembles into a trimer. Each F protein subunit is

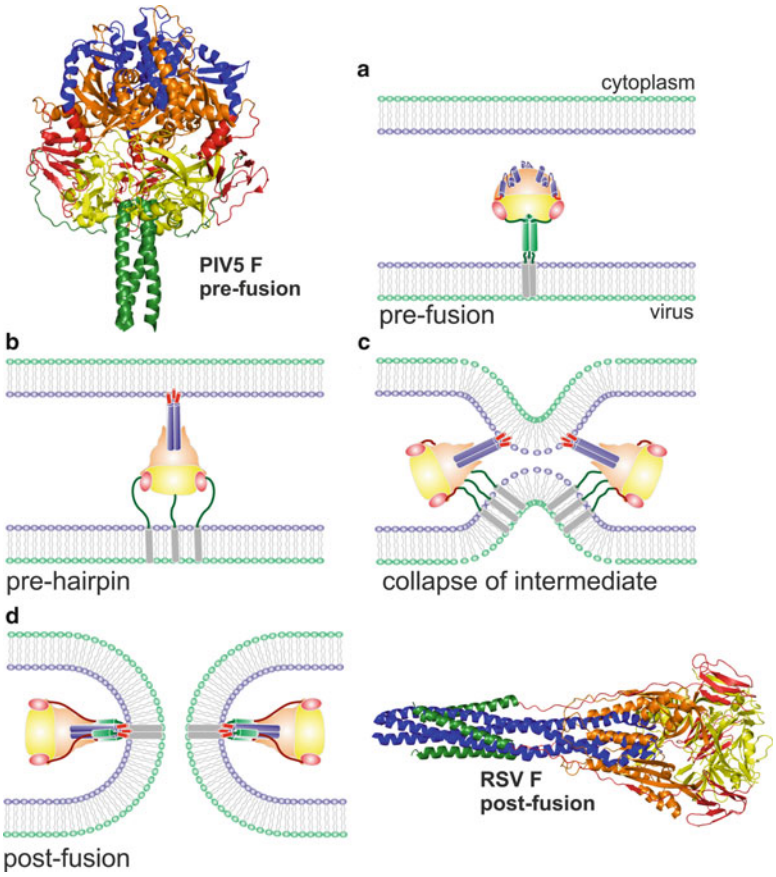


Fig. 16.2 Membrane fusion mediated by a class I fusion protein (Paramyxovirus). The atomic structures of the pre-fusion form of Parainfluenza virus type 5 (PIV5) [22] (*upper left*) and the post-fusion form of Respiratory Syncytial Virus (RSV) [53] (*lower right*) F proteins are shown as ribbons. The same protein regions are highlighted with identical colors in the two conformations. **(a–d)** Diagram of the fusion process denoting: **(a)** the pre-fusion paramyxovirus F protein trimer inserted in the viral membrane before activation, **(b)** formation of the pre-hairpin structure which includes refolding of the long central HRA α-helices (*blue*) with the fusion peptide (*red*) inserted into the cell membrane, **(c)** collapse of the pre-hairpin to approach the two membranes, and **(d)** formation of the fusion pore and stabilization of the F trimer in the post-fusion conformation. In the last two steps, two F protein molecules are represented to indicate the cooperation needed to drive the fusion process

proteolytically cleaved during transport to the cell surface, generating two chains, F2 N-terminal and F1 C-terminal that remain linked by one or more disulfide bonds. F1 (equivalent to the HA2 chain of influenza virus) has a hydrophobic fusion peptide at the N-terminus and two heptad repeat sequences (HRA and HRB) in its ectodomain. HRA is adjacent to the fusion peptide and HRB is proximal to the transmembrane (TM) region, which is placed near the F1 C-terminus.

The pre-fusion three-dimensional (3D) structure of the parainfluenza virus type 5 (PIV5) F protein contains a large globular head connected to a short trimeric coiled-coil made by the HRB region [22] (Fig. 16.2). Comparison with the post-fusion structure of the F ectodomain from other paramyxovirus (for instance respiratory syncytial virus (RSV (Fig. 16.2)) and functional studies using peptide inhibitors [24] provided the following model of membrane fusion: Upon virus binding to cells through the attachment protein, F is activated and initiates a series of conformational changes, including separation of the HRB coils and refolding of HRA sequences to form a very elongated trimeric coiled-coil. The fusion peptides -now at the N-terminus of the HRAs- insert into the target cell membrane, resulting in formation of the pre-hairpin intermediate. This step is followed by zipping of the C-terminal part of the molecule along the core coiled-coil to bring together the two membranes and the fusion and TM domains, in analogy with the process described before for influenza HA. However, in the case of the paramyxovirus F the HRB sequences wrap around the HRA coiled-coil forming an extremely stable six-helix bundle (6HB) in the post-fusion hairpin. Formation of this 6HB provides most of the energy required to overcome membrane repulsion. The 6HB structure is shared by other class I fusion glycoproteins, such as the gp41 chain of the HIV envelope glycoprotein.

While activation of influenza HA requires exposure to the endosomal low pH (probably by protonation of key amino acid residues), the event that triggers paramyxovirus F proteins is still ill-defined. Cell-cell fusion of transfected cells that express paramyxovirus F requires in most cases co-expression of the homotypic attachment protein, suggesting that an interaction of the two proteins is needed for membrane fusion. Two alternative models ("clamp" and "provocateur") have been proposed to explain the requirement of the attachment protein for fusion:

1. The *clamp* model postulates that HN (or the equivalent attachment protein depending on the virus) is complexed with F in the virus particle, retaining the latter in the metastable configuration. Conformational changes in HN upon receptor binding release F from the complex to initiate membrane fusion.
2. Alternatively, the *provocateur* model postulates that HN and F do not interact in the virus before contacting the cell. Concomitantly to the structural changes induced in HN upon receptor binding, HN binds to F and this interaction triggers F for fusion [25].

Intriguingly, the F protein of viruses belonging to the *Pneumovirinae* subfamily of paramyxoviruses (*e.g.*, RSV and human metapneumovirus, hMPV) do not require co-expression of the attachment protein (G) for cell-cell fusion [26]. Furthermore, deletion mutant viruses have been obtained in which the entire G gene is obliterated. These mutants still infect cells *in vitro*, although less efficiently than the wild type virus and are attenuated in animal models of infection [27]. Activation of the F protein of those deletion mutants cannot rely on interactions with the G protein and therefore alternative regulatory mechanisms should control membrane fusion. Of note, a unique characteristic of the RSV F protein is the presence of two proteolytic cleavage sites (instead of one, as in all other

paramyxovirus) in the F0 protein precursor [26]. The presence of a double cleavage site in F has been found to influence membrane fusion activation by a still poorly understood G independent mechanism [28].

16.3.2 Class II Viral Fusion Proteins

In contrast to class I fusion proteins, the so-called class II fusion proteins (Table 16.1) are derived from a polyprotein precursor that is cleaved during biosynthesis to generate the E1 protein of alphaviruses (*e.g.*, Semliki Forest virus (SFV)) or the E protein of flaviviruses (*e.g.*, dengue virus and tick-borne encephalitis virus). Both proteins fold co-translationally with a companion or regulatory protein, termed p62 for alphaviruses and prM for flaviviruses [9].

In alphavirus, the p62-E1 complex is transported to the plasma membrane where they are incorporated into new budding icosahedral virus particles as dimers of p62-E1. p62 is then proteolitically processed (and then named E2) but remains bound to the virus where it covers most of the fusion protein E1 and specially its fusion loop. E2 mediates binding to the cell surface receptor.

In contrast, flavivirus particles bud into the endoplasmic reticulum as immature virions containing prM-E protein complexes. The immature viruses are then transported to the exterior through the exocytic pathway where prM is processed and separated from E [29]. The latter protein is then arranged in E-E homodimers at the virion surface with icosahedral symmetry. The flavivirus fusion E protein is additionally responsible for receptor binding.

The first structure of any class II glycoprotein, solved by X-ray crystallography, was that of the tick-borne encephalitis (TBE) flavivirus E protein ectodomain [30], solubilized from virions by limited trypsin digestion. Similar structures have now been solved for the E ectodomain of dengue virus types 2 and 3 [9]. The polypeptide chain of the E protein follows a complex path, resulting in three globular domains, essentially constituted by β-sheets (Fig. 16.3). The first domain is a β-barrel with up-and-down topology (red). Two adjacent strands in domain I are extended, forming domain II (yellow) which is a long "finger-like" structure that runs parallel to the viral membrane. At the tip of domain II is the hydrophobic fusion loop which remains buried in the virion from the hydrophilic environment by interaction with domain III (blue) of the adjacent monomer in the E-E dimer. Domain I is also connected to domain III which bridges the E ectodomain with the so-called stem region that extends to the TM region of the protein.

Unlike the class I fusion proteins, which are trimeric in their pre- and post-fusion conformations, class II fusion glycoproteins undergo major oligomeric transformations during fusion. As in the case of influenza virus, the flavivirus E protein first binds to a cell surface receptor which induces endocytosis of the virion. Once in the acidic endosome, the E-E homodimer dissociates, resulting in disassembly of the icosahedral scaffold. The individual subunits swing outward by the hinge region that connects domains I and II, and the fusion loops insert into the target membrane. Lateral

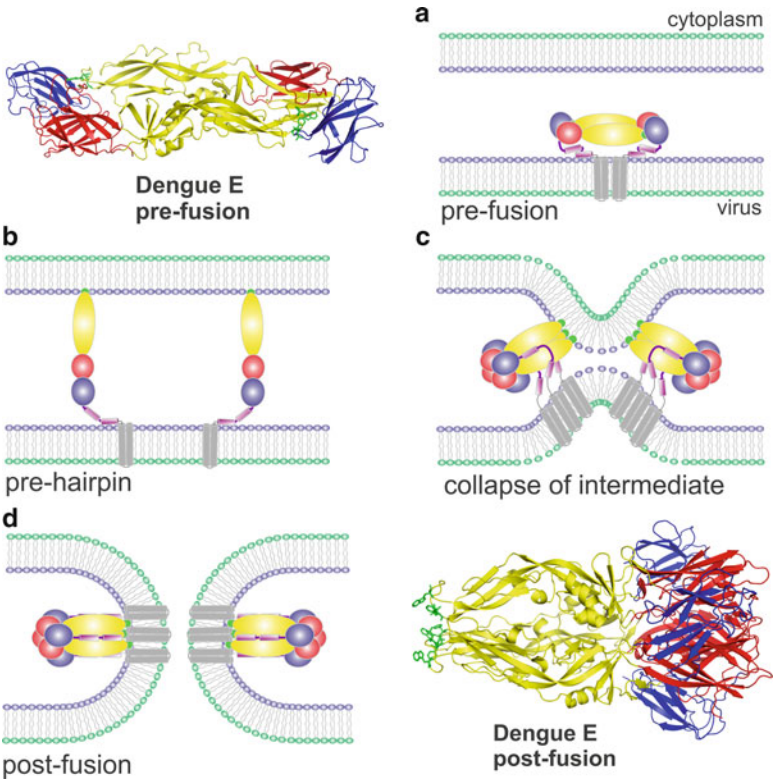


Fig. 16.3 Membrane fusion mediated by a class II fusion protein (Flavivirus). Ribbon representation of the atomic structures of the dengue virus E protein dimer in the pre-fusion conformation [54] (*upper left*) and the E protein trimer in the post-fusion conformation [55] (*lower right*). Domains I, II and III of the E glycoprotein are colored *red*, *yellow* and *blue*, respectively. **(a–d)** Diagram of the fusion process denoting: **(a)** the structure of the flavivirus E glycoprotein dimer already in the endosome before activation, **(b)** dissociation of the E protein subunits, refolding of the fusion domain (*yellow*) and insertion of the fusion loop (*green*) into the endosomal membrane, **(c)** formation and refolding of the E protein trimer to approach the two membranes, and **(d)** formation of the fusion pore and stabilization of the E trimer in the post-fusion conformation. In the last two steps, two E protein molecules are represented to indicate the cooperation needed to drive the fusion process

interactions between monomers facilitates reclustering into trimers [31]. These rearrangements lead to the formation of an extended trimeric structure, analogous to the pre-hairpin intermediate of class I fusion glycoproteins, in which two different regions of each E polypeptide are inserted into the two membranes to be fused. Collapse of the extended intermediate can proceed by rotation of domain III in each subunit about the segment that links it to domain I and zipping up of the stem alongside the clustered domains II. This refolding brings the two membranes together to initiate formation of the lipid stalk, the hemifusion diaphragm and the fusion pore.

The structure of the fusion E1 glycoprotein of alphaviruses (SFV) was found unexpectedly very similar to that of the flavivirus E protein, despite the lack of

detectable sequence conservation. E1 has also three discernible domains, equivalent to those of flavivirus E. The only significant difference is the association of E1 with E2 in the virus particle. E2 interacts with the cell surface receptor to initiate the endocytic internalization of the SFV virion [32]. In the acidic endosome, E2 separates from E1 and it is probably degraded. Upon low pH exposure, E1 undergoes similar conformational changes to those of the flavivirus E protein, leading to fusion of the viral and endosomal membranes. Electron microscopy and X-ray crystallography results provide support for interactions between adjacent E1 trimers when the fusion loops are inserted in the target membrane to produce rings of five or six trimers. It has been postulated that these fivefold interactions would act at the fusion site to induce the formation of a nipple-like curvature in the viral and target membranes, favoring membrane fusion [33]. Although there is no direct evidence, it is likely that the flavivirus E protein forms similar rings of trimers during fusion.

16.3.3 Class III Viral Fusion Proteins

The best characterized members of the so-called class III fusion viral glycoproteins are the rhabdovirus (*e.g.*, vesicular stomatitis virus, VSV) G glycoprotein, the herpesvirus gB glycoprotein and the baculovirus gp64 glycoprotein. The pre-fusion [34] and post-fusion [35] structures of the VSV_G glycoprotein ectodomain have been solved by X-ray crystallography while only the post-fusion conformations of gB [36] and gp64 [37] are known.

Class III fusion glycoproteins are expressed from individual mRNAs and do not require proteolytic processing of either a protein precursor (as in class I proteins) or an accompanying protein (as in class II proteins) for activity. Class III proteins are trimeric before and after fusion and share structural characteristics with both class I and class II fusion glycoproteins, as described below.

The rhabdovirus G protein possesses both receptor binding and fusion promoting activities. As in the case of influenza virus, binding of rhabdovirus G to a poorly characterized receptor at the cell surface induces endocytosis of the virus particle. Acidification of the endosome triggers G for membrane fusion. However, and in contrast with all other fusion proteins, the low pH inactivation of rhabdovirus G is reversible. Thus, virions inactivated by prolonged incubation at pH <6 can be reactivated by raising the pH to neutral [38]. This reversibility may be required to allow G to be transported through the acidic Golgi apparatus and to recover its native fusion-competent state when incorporated to new virions [39]. Given this reversibility, it is believed that the energy released during the structural transition of a single trimer from the pre-fusion to the post-fusion conformation is probably small, compared with the energetic barrier of the fusion reaction. In agreement with this hypothesis, the estimated number of rhabdovirus spikes required for fusion is higher (at least 15 trimers) than for other enveloped viruses.

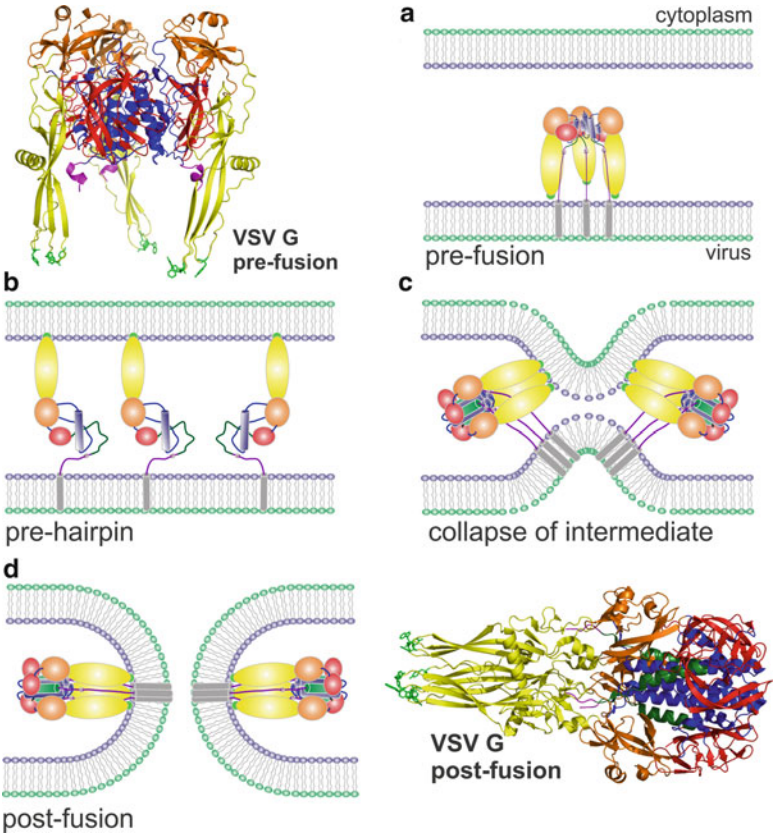


Fig. 16.4 Membrane fusion mediated by a class III fusion protein (Vesicular Stomatitis Virus, VSV). Ribbon representations of the VSV glycoprotein (G), in the pre- (*upper left*) and post-fusion (*lower right*) conformations. Domains are colored similarly in all images. The fusion domain is colored in *yellow* and the fusion loops in *green*. **(a–d)** Diagram of the fusion process denoting: **(a)** the structure of the VSV G glycoprotein trimer already in the endosome before activation, **(b)** dissociation of the G protein subunits, refolding of the fusion domain (*yellow*) and insertion of the fusion loop (*green*) into the endosomal membrane, **(c)** formation and refolding of the G protein trimer to approach the two membranes, and **(d)** formation of the fusion pore and stabilization of the G primer in the post-fusion conformation. In the last two steps, two G protein molecules are represented to indicate the cooperation needed to drive the fusion process

A soluble ectodomain of the VSV_G glycoprotein, released from purified virions treated with thermolysin, was used to solve the structures of the pre-fusion and post-fusion conformations, after exposure to high and low pH, respectively [34, 35]. Several domains could be observed in both structures that are rearranged in their relative orientations during transit from the pre- to the post-fusion structure (Fig. 16.4). In the pre-fusion conformation, the fusion domain contains two fusion loops reminiscent of class II proteins that are oriented downward towards the viral membrane. After low pH exposure, the fusion domain moves upward by flipping

relative to the central core of the trimer. Thus, an intermediate equivalent to the pre-hairpin structure of class I proteins is formed. This is followed by the reversal of the molecule around a central rigid block formed by lengthening of the central helix and refolding of the three C-terminal segments into helices that position themselves in the grooves of the central core in an anti-parallel manner. This six-helix bundle has obvious resemblance with that of the class I proteins.

It is likely that the transition of VSV_G from the pre-fusion to the post-fusion conformation involves disassembly of the trimer into monomers and reassembly into trimers upon interaction of the fusion loops with the target membrane [40]. It is also likely that cooperativity between G glycoproteins is needed to overcome the energy barrier, as mentioned above. As for class II glycoproteins, lattices of G proteins have been observed in virions, particularly in the planar base of the rhabdovirus bullet-shape particle, which may act to induce nipple-like deformations in the viral and target membranes.

Although it has not been reported, it is likely that low pH exposure also leads to reversible inactivation of the baculovirus gp64 glycoprotein. This protein, like VSV_G, is involved in both receptor binding and membrane fusion after endosome acidification, since baculoviruses also use an endocytic route of entry [41]. In contrast, membrane fusion mediated by herpesvirus gB can occur either at the plasma membrane or in endosomes, depending on the virus and the target cell type. In either case, attachment of herpesvirus to host cells follows a complex mechanism in which several viral glycoproteins interact with cell surface molecules. Some of these interactions trigger fusion, whereas others simply serve to tether the virus to the cell and are dispensable for fusion. In any case, the gB protein, shared by all viruses of the *Herpesviridae* family, is responsible for fusion [42].

16.3.4 Other Viral Fusion Proteins

Poxviruses (vaccinia virus is the best known member) represent an extreme case among enveloped viruses, regarding the number of viral glycoproteins required for entry. As for herpes virus, entry can occur by fusion at the plasma membrane or in a low pH-dependent manner from within an intracellular particle, depending on the virus strain and the cell type. Vaccinia virus internalization is believed to occur by macropinocytosis (a type of non-specific endocytosis). Four vaccinia virus proteins are involved in attachment to cell surface proteoglycans or laminin [43]. Eleven or 12 other relatively small glycoproteins, ranging in size between 35 and 377 amino acids, form the so-called entry fusion complex (EFC) that mediates membrane fusion [44]. These proteins have N- or C-terminal transmembrane domains but no sequence similarity with the fusion peptide of other viral fusion glycoproteins has been found in any of them. Therefore, the actual mechanism of vaccinia virus membrane fusion remains to be elucidated but it seems to be different from that of other enveloped viruses. By using conditional lethal mutants of each of the 11 proteins that make the EFC, it was found that eight of them were required to reach

the hemifusion step and the other three were needed for completion of virus entry [44]. It is likely that hydrophobic regions of several proteins may assemble in the EFC to form a hydrophobic surface that could bind to the target membrane and drive membrane fusion by some novel mechanism.

Finally, the fusion-associated small transmembrane (FAST) proteins of reoviruses are brought here -despite not being involved in virus entry and reovirus being a non-enveloped virus- because they induce cell-cell fusion and therefore facilitates dissemination of virus to neighboring cells. The FAST proteins are small non structural proteins (98–148 amino acids, depending on the viral strain) that are expressed on the surfaces of virus-infected cells, where they induce cell-cell fusion and syncytia (multinucleate cells) formation. Purified FAST proteins, when reconstituted into liposome membranes, induce fusion indicating that they are *bona fide* fusogens [45]. The orientation of the FAST polypeptides in the cell membrane is also unique among viral fusogens, with a relatively short N-terminal ectodomain followed by a transmembrane region and a long C-terminal cytoplasmic tail. Although they lack a fusion peptide, a relatively hydrophobic region near the N-terminus which is additionally myristoylated seems to insert into the target membrane to drive membrane fusion, at least for certain FAST proteins [46] .

16.4 Early Post-Entry Events

Once membrane fusion has been completed, the viral genome -generally in complex with other proteins or inside a viral nucleocapsid (see Chaps. 2 and 11)- is found for a second time in a cytoplasmic environment. The first time is when the genome assembles in the cytosol of the infected cell or when it is trafficking from the nucleus to the cell exterior, depending on the virus. However, the fate of the incoming genome is now very different and characteristic for each virus.

Most RNA viruses replicate in the cell cytoplasm, although there are exceptions like influenza virus or borna virus that do so in the nucleus. If the RNA is of positive polarity, like in flavivirus, the genome may act as mRNA to be translated by the cell protein synthesis machinery. In most cases the primary translational product is a polyprotein that matures into the different viral gene products by proteolytic processing [47]. In the case of negative-stranded RNA viruses, like paramyxoviruses, the first step after entry is the transcription of the viral genome to yield the different mRNAs that are translated into the distinct viral gene products [48].

For RNA viruses that replicate in the nucleus, the nucleoprotein complex of the viral genome and associated proteins has to be transported to the cell nucleus for transcription. Most of the viral proteins required in the nucleus have their own nuclear localization signal (NLS), which is a cluster of basic amino acids. However, actual import of the viral ribonucleoprotein into the cell nucleus may require additional interactions with certain host factors. For instance, the NLS of the influenza nucleoprotein interacts with karyopherin α and this in turn with karyopherin β which

mediates interactions with the proteins of the nuclear pore to promote nuclear import of the viral ribonucleoproteins by an energy-dependent process [49].

Retroviruses represent a special case in which a RNA viral genome (diploid), packed in a capsid inside the virus envelope has to be transcribed to DNA before integration into the cell host genome. In this case, fusion of the virus and cell membrane delivers the capsid into the cell cytoplasm, where it interacts with cytoskeleton and other cell components for transport to the vicinity of the nucleus where reverse transcription and uncoating takes place. Then, the resulting pre-integration complexes are transported through the nuclear pore inside the nucleus for integration into the host genome [50].

Most DNA viruses replicate in the nucleus, with exceptions like poxviruses (*e.g.*, vaccinia virus). In the case of herpesviruses, the capsids that contain the viral genome are transported from the site of release to the nuclear pore where importin β promotes nuclear import of the viral DNA by an energy-dependent mechanism [51]. In contrast, the virus core of vaccinia virus is released into the cytoplasm. The virus core has all the machinery for transcription of the early genes that ensues further progress on DNA replication and transcription of the remaining genome [52].

16.5 Perspectives and Conclusions

As noted, viral-cell membrane fusion is the last extracellular step of enveloped viruses and therefore still amenable to inhibition by chemical or biological products without the drawback of the membrane permeability barrier. Hence, it is not surprising that viral fusogens have received recently very much attention as ideal targets for the development of effective antivirals, some of them already in clinical use (*e.g.*, the T20 peptide for HIV or the Synagis antibody for respiratory syncytial virus) (see also Chap. 20). The development of high throughput technologies for screening of large libraries of chemical or biological molecules (*e.g.*, antibodies) should provide in the near future a plethora of drugs to fight some of the important diseases caused by enveloped virus.

A critical step in the process of viral induced membrane fusion is the activation of the viral fusogen. While most viruses that enter the cell through low pH endocytosis rely on protonation of certain residues (mostly histidines) of the viral fusogen to trigger fusion, the activating step in the case of other viruses is still ill-defined. Better understanding of fusion triggering may therefore bring new possibilities for the manipulation of virus entry. Finally, determination of the structures of certain fusion intermediates could provide additional targets for antivirals.

Learning how virus-cell membrane fusion proceeds may also be relevant for other spheres of Biology. For instance, regulation of vesicle fusion could benefit from knowledge acquired in the field of Virology and find applications in studies of synaptic transmission and Neurobiology.

Acknowledgements Current research activities in the Biología Viral laboratory are funded by grants GR09/0039 from Instituto de Salud Carlos III and SAF2009-11632 from Plan Nacional de I + D + i.

References and Further Reading

1. Chernomordik LV, Kozlov MM (2008) Mechanics of membrane fusion. Nat Struct Mol Biol 15:675–683
2. Yang L, Huang HW (2002) Observation of a membrane fusion intermediate structure. Science 297:1877–1879
3. Chanturiya A, Chernomordik LV, Zimmerberg J (1997) Flickering fusion pores comparable with initial exocytotic pores occur in protein-free phospholipid bilayers. Proc Natl Acad Sci U S A 94:14423–14428
4. Harrison SC (2008) Viral membrane fusion. Nat Struct Mol Biol 15:690–698
5. Munoz-Barroso I, Durell S, Sakaguchi K, Appella E, Blumenthal R (1998) Dilation of the human immunodeficiency virus-1 envelope glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41. J Cell Biol 140:315–323
6. Floyd DL, Ragains JR, Skehel JJ, Harrison SC, van Oijen AM (2008) Single-particle kinetics of influenza virus membrane fusion. Proc Natl Acad Sci U S A 105:15382–15387
7. Zimmerberg J, Blumenthal R, Sarkar DP, Curran M, Morris SJ (1994) Restricted movement of lipid and aqueous dyes through pores formed by influenza hemagglutinin during cell fusion. J Cell Biol 127:1885–1894
8. Danieli T, Pelletier SL, Henis YI, White JM (1996) Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. J Cell Biol 133:559–569
9. Kielian M, Rey FA (2006) Virus membrane-fusion proteins: more than one way to make a hairpin. Nat Rev Microbiol 4:67–76
10. Yang X, Kurteva S, Ren X, Lee S, Sodroski J (2006) Subunit stoichiometry of human immunodeficiency virus type 1 envelope glycoprotein trimers during virus entry into host cells. J Virol 80:4388–4395
11. Magnus C, Rusert P, Bonhoeffer S, Trkola A, Regoes RR (2009) Estimating the stoichiometry of human immunodeficiency virus entry. J Virol 83:1523–1531
12. Chen J, Skehel JJ, Wiley DC (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. Proc Natl Acad Sci U S A 96:8967–8972
13. Buzon V, Natrajan G, Schibli D, Campelo F, Kozlov MM, Weissenhorn W (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. PLoS Pathog 6:e1000880
14. Kemble GW, Danieli T, White JM (1994) Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. Cell 76:383–391
15. Chernomordik LV, Kozlov MM (2003) Protein-lipid interplay in fusion and fission of biological membranes. Annu Rev Biochem 72:175–207
16. Fasshauer D, Otto H, Eliason WK, Jahn R, Brunger AT (1997) Structural changes are associated with soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor complex formation. J Biol Chem 272:28036–28041
17. Rickman C, Hu K, Carroll J, Davletov B (2005) Self-assembly of SNARE fusion proteins into star-shaped oligomers. Biochem J 388:75–79
18. Rizo J, Rosenmund C (2008) Synaptic vesicle fusion. Nat Struct Mol Biol 15:665–674
19. Sollner T, Bennett MK, Whiteheart SW, Scheller RH, Rothman JE (1993) A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. Cell 75:409–418

20. Mohler WA, Shemer G, del Campo JJ, Valansi C, Opoku-Serebuoh E, Scranton V, Assaf N, White JG, Podbilewicz B (2002) The type I membrane protein EFF-1 is essential for developmental cell fusion. Dev Cell 2:355–362
21. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 289:366–373
22. Yin HS, Wen X, Paterson RG, Lamb RA, Jardetzky TS (2006) Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. Nature 439:38–44
23. Yin HS, Paterson RG, Wen X, Lamb RA, Jardetzky TS (2005) Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. Proc Natl Acad Sci USA 102:9288–9293
24. Russell CJ, Jardetzky TS, Lamb RA (2001) Membrane fusion machines of paramyxoviruses: capture of intermediates of fusion. EMBO J 20:4024–4034
25. Connolly SA, Leser GP, Jardetzky TS, Lamb RA (2009) Bimolecular complementation of paramyxovirus fusion and hemagglutinin-neuraminidase proteins enhances fusion: implications for the mechanism of fusion triggering. J Virol 83:10857–10868
26. Gonzalez-Reyes L, Ruiz-Arguello MB, Garcia-Barreno B, Calder L, Lopez JA, Albar JP, Skehel JJ, Wiley DC, Melero JA (2001) Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. Proc Natl Acad Sci U S A 98:9859–9864
27. Biacchesi S, Skiadopoulos MH, Yang L, Lamirande EW, Tran KC, Murphy BR, Collins PL, Buchholz UJ (2004) Recombinant human Metapneumovirus lacking the small hydrophobic SH and/or attachment G glycoprotein: deletion of G yields a promising vaccine candidate. J Virol 78:12877–12887
28. Rawling J, Cano O, Garcin D, Kolakofsky D, Melero JA (2011) Recombinant sendai viruses expressing fusion proteins with two furin cleavage sites mimic the syncytial and receptor-independent infection properties of respiratory syncytial virus. J Virol 85:2771–2780
29. Li L, Lok SM, Yu IM, Zhang Y, Kuhn RJ, Chen J, Rossmann MG (2008) The flavivirus precursor membrane-envelope protein complex: structure and maturation. Science 319:1830–1834
30. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375:291–298
31. Bressanelli S, Stiasny K, Allison SL, Stura EA, Duquerroy S, Lescar J, Heinz FX, Rey FA (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. EMBO J 23:728–738
32. Lescar J, Roussel A, Wien MW, Navaza J, Fuller SD, Wengler G, Wengler G, Rey FA (2001) The fusion glycoprotein shell of Semliki forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell 105:137–148
33. Gibbons DL, Vaney MC, Roussel A, Vigouroux A, Reilly B, Lepault J, Kielian M, Rey FA (2004) Conformational change and protein-protein interactions of the fusion protein of Semliki forest virus. Nature 427:320–325
34. Roche S, Rey FA, Gaudin Y, Bressanelli S (2007) Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science 315:843–848
35. Roche S, Bressanelli S, Rey FA, Gaudin Y (2006) Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science 313:187–191
36. Backovic M, Longnecker R, Jardetzky TS (2009) Structure of a trimeric variant of the Epstein-Barr virus glycoprotein B. Proc Natl Acad Sci U S A 106:2880–2885
37. Kadlec J, Loureiro S, Abrescia NG, Stuart DI, Jones IM (2008) The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. Nat Struct Mol Biol 15:1024–1030
38. Roche S, Gaudin Y (2002) Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers. Virology 297:128–135
39. Albertini AA, Baquero E, Ferlin A, Gaudin Y (2012) Molecular and cellular aspects of rhabdovirus entry. Viruses 4:117–139

40. Albertini AA, Merigoux C, Libersou S, Madiona K, Bressanelli S, Roche S, Lepault J, Melki R, Vachette P, Gaudin Y (2012) Characterization of monomeric intermediates during VSV glycoprotein structural transition. PLoS Pathog 8:e1002556
41. Hefferon KL, Oomens AG, Monsma SA, Finnerty CM, Blissard GW (1999) Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. Virology 258:455–468
42. Connolly SA, Jackson JO, Jardetzky TS, Longnecker R (2011) Fusing structure and function: a structural view of the herpesvirus entry machinery. Nat Rev Microbiol 9:369–381
43. Ho Y, Hsiao JC, Yang MH, Chung CS, Peng YC, Lin TH, Chang W, Tzou DL (2005) The oligomeric structure of vaccinia viral envelope protein A27L is essential for binding to heparin and heparan sulfates on cell surfaces: a structural and functional approach using site-specific mutagenesis. J Mol Biol 349:1060–1071
44. Laliberte JP, Weisberg AS, Moss B (2011) The membrane fusion step of vaccinia virus entry is cooperatively mediated by multiple viral proteins and host cell components. PLoS Pathog 7: e1002446
45. Top D, de Antueno R, Salsman J, Corcoran J, Mader J, Hoskin D, Touhami A, Jericho MH, Duncan R (2005) Liposome reconstitution of a minimal protein-mediated membrane fusion machine. EMBO J 24:2980–2988
46. Top D, Read JA, Dawe SJ, Syvitski RT, Duncan R (2012) Cell-cell membrane fusion induced by p15 fusion-associated small transmembrane (FAST) protein requires a novel fusion peptide motif containing a myristoylated polyproline type II helix. J Biol Chem 287:3403–3414
47. Lindenbach BD, Thiel H-J, Rice CM (2012) Flaviviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields Virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia
48. Lamb RA, Parks GD (2012) Paramyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia
49. Melen K, Fagerlund R, Franke J, Kohler M, Kinnunen L, Julkunen I (2003) Importin alpha nuclear localization signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. J Biol Chem 278:28193–28200
50. Arhel N (2010) Revisiting HIV-1 uncoating. Retrovirology 7:96
51. Roizman B, Knipe DM, Whitley RJ (2012) Herpes simplex viruses. In: Knipe DM, Howley PM (eds) Fields virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia
52. Broyles SS (2003) Vaccinia virus transcription. J Gen Virol 84:2293–2303
53. McLellan JS, Yang Y, Graham BS, Kwong PD (2011) Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. J Virol 85:7788–7796
54. Modis Y, Ogata S, Clements D, Harrison SC (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc Natl Acad Sci U S A 100:6986–6991
55. Modis Y, Ogata S, Clements D, Harrison SC (2004) Structure of the dengue virus envelope protein after membrane fusion. Nature 427:313–319

Further Reading

Palfreyman MT, Jorgensen EM (2009) *In vivo* analysis of membrane fusion. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Chichester. doi:10.1002/9780470015902.a0020891

Grove J, Marsh M (2011) The cell biology of receptor-mediated virus entry. J Cell Biol 7:1071–1082

Helenius A (2007) Virus entry and uncoating. In: Knipe DM, Howley PM (eds) Fields virology, 5th edn. Lippincott, Williams and Wilkins, Philadelphia

Also especially recommended for further reading are references [4, 9, 15, 42, 44] *listed above*

Chapter 17
Bacteriophage Receptor Recognition and Nucleic Acid Transfer

Carmela Garcia-Doval and Mark J. van Raaij

Abstract Correct host cell recognition is important in the replication cycle for any virus, including bacterial viruses. This essential step should occur before the bacteriophage commits to transfer its genomic material into the host. In this chapter we will discuss the proteins and mechanisms bacteriophages use for receptor recognition (just before full commitment to infection) and nucleic acid injection, which occurs just after commitment. Some bacteriophages use proteins of the capsid proper for host cell recognition, others use specialised spikes or fibres. Usually, several identical recognition events take place, and the information that a suitable host cell has been encountered is somehow transferred to the part of the bacteriophage capsid involved in nucleic acid transfer. The main part of the capsids of bacteriophages stay on the cell surface after transferring their genome, although a few specialised proteins move with the DNA, either forming a conduit, protecting the nucleic acids after transfer and/or functioning in the process of transcription and translation.

Keywords Adsorption • Attachment • Bacterium • Bacteriophage • Baseplate • Capsid • Cell wall • Flagellum • Nucleic acid conduit • Contractile tail • Gram-negative • Gram-positive • Lipopolysaccharide • Membrane • Peptidoglycan • Pilus • Receptor • Tail fibre • Tailspike • Teichoic acid

C. Garcia-Doval • M.J. van Raaij (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnología (CSIC), c/Darwin 3, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: mjvanraaij@cnb.csic.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_17,

Abbreviations

DNA	Deoxyribonucleic acid
EMDB	Electron microscopy data bank
gp	Gene product
Hoc	Highly antigenic outer capsid protein
LPS	Lipo-polysaccharide
nm	Nanometer
Omp	Outer membrane porin
ORF	Open reading frame
PDB	Protein structure data bank
RNA	Ribonucleic acid
Soc	Small outer capsid protein
stf	Side tail fibre
tfa	Tail fibre assembly

17.1 Introduction

Bacteriophages, or phages, are viruses that infect bacteria and were independently discovered by both Félix d'Herelle and Frederick Twort. Bacteriophages are now known to be the most numerous replicating biological entities on earth; for every bacterium there are thought to be on average ten phages [1]. The observation that different bacteriophages specifically recognise and kill different bacterial species and strains has led to multiple applications [2]. Félix d'Herelle himself pioneered phage therapy, although with the advent of modern antibiotics it is not commonly used anymore. This situation may change now as more and more bacterial strains develop resistance against many or all useful antibiotics. Bacteriophages and the lytic enzymes they produce are also investigated for the control of microbes in the food industry, of water-borne pathogens and pathogenic bacteria in hospital settings. A much-used application of bacteriophages is phage typing, in which a bacterial species can often be identified at the strain level based on its susceptibility to a library of different bacteriophages.

Phages may interfere with industrial fermentation processes such as those in dairy plants and have important negative economic consequences. Bacteriophages, especially lysogenic ones, may also transfer DNA encoding virulence factors to bacteria, making them more pathogenic. Therefore, bacteriophages to be used in the food industry and medical applications should be well characterised and proven to be free of sequences encoding virulence factors.

In the laboratory, phages have been used in many fundamental experiments in microbiology, genetics, biochemistry and molecular biology since Max Delbrück

and Salvador Luria started investigating them in the middle of the twentieth century. Nowadays, bacteriophages are used as cloning and phage display vectors, DNA encoding phage T7 RNA polymerase is part of protein expression vectors [3] and phage enzymes such as DNA polymerases and ligases are used in routine DNA manipulations. The ease with which many bacteriophages can be grown, their stability and relatively facile genetic modification has led to the development of advanced applications such as their use as labelled ground tracers to follow the flow of underground water, as innocuous test particles for filters designed to remove dangerous viruses and phage-based diagnostics to detect low amounts of dangerous bacteria [2]. Some phages are being used as nanoparticles for nanotechnological developments (see Chap. 22).

Bacteriophages can be divided into virulent, or obligatory lytic, phages on one hand and temperate, or lysogenic phages on the other. Upon infection of their host cell, temperate phages may take one of two paths, lytic or lysogenic. In the lytic pathway, the phage takes over the host cell and redirects its biosynthetic pathways towards producing many daughter phages and ultimately, lysis of the host cell to release its progeny. Alternatively, a temperate phage may integrate its DNA into the host genome – this phage form is called a prophage. In this way, its genome replicates with that of the bacterium in conditions favourable for bacterial growth (some phages even encode virulence factors that favour bacterial growth inside a eukaryotic host). When conditions cease to be favourable for bacterial growth and the bacterial SOS system is activated, the phage may switch to the lytic pathway, produce daughter phages, lyse the bacterium and "escape". Virulent phages do not have the ability to integrate into the host cell genome and can only follow the lytic pathway. Filamentous phages (see below) are an exception and follow an intermediate pathway. When they infect bacteria, they redirect the bacterial biosynthetic systems to produce phages, but phage egress does not lead to lysis. Instead, the bacterium turns into a phage factory, continuously extruding phages. The bacterium continues to grow and divide, albeit at a slower rate.

Bacteriophages can be divided into several different types based on their size and nature of their genetic material: large DNA phages, small DNA phages and RNA phages. Large DNA phages have a double-stranded DNA genome of 20 kilobases or more and can infect both Gram-negative and Gram-positive bacteria. Probably all large DNA phages belong to the *Caudovirales* order, as no phages with genomes greater than 15 kilobases have been found without a tail. Their characteristic tails function in efficient host cell recognition and infection and can be long, straight and contractile (*Myoviridae* family), long, flexible and non-contractile (*Siphoviridae* family), or short and stubby (*Podoviridae* family). Empty phage particles stay on the cell surface (unlike most other viruses). Well-known examples of myoviruses are the virulent *Escherichia coli* phage T4 (Fig. 17.1a) and the temperate *Escherichia coli* phage μ, while examples of *Escherichia coli* siphoviruses are the virulent phage T5 and the temperate phage λ (Fig. 17.1b). Examples of podoviruses are the virulent *Escherichia coli* phage T7, the temperate *Salmonella* phage P22 (Fig. 17.1c) and the virulent *Bacillus subtilis* phage ϕ29. Phages with long tails invariably contain a tape

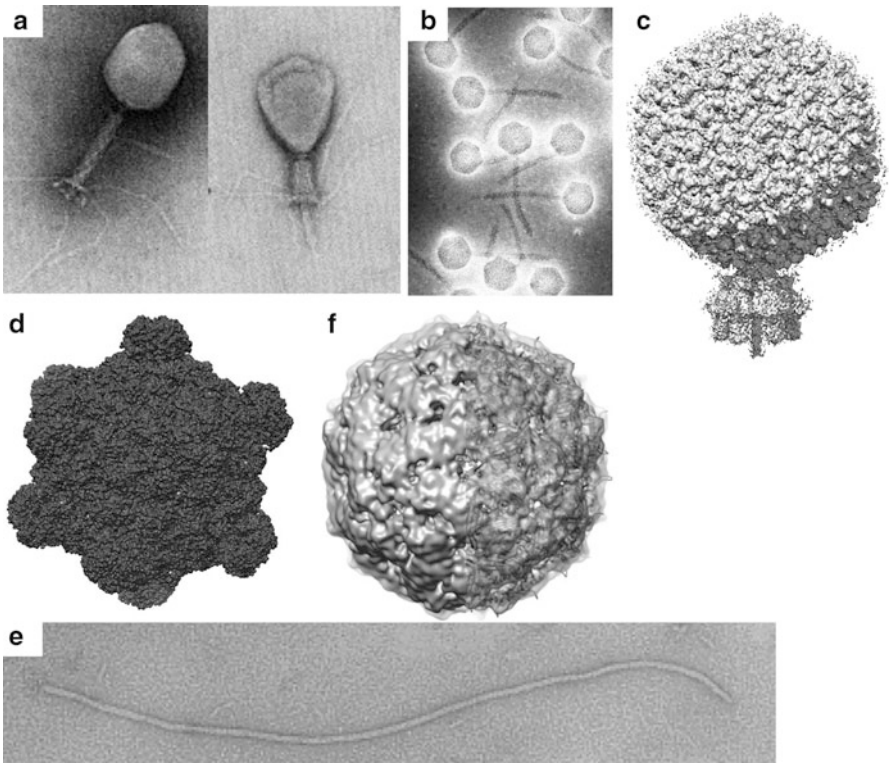


Fig. 17.1 Diverse members of the bacteriophage family. (**a**) Electron micrographs of extended (*left*) and contracted (*right*) T4 phage; the capsid is about 86 nm wide (Reproduced from [4]. With permission). (**b**) Electron micrograph of bacteriophage λ; the capsid width is 550 nm (Reproduced from [5]. With permission). (**c**) Cryo-EM density of phage P22 [6]; here the capsid has a width of 70 nm (figure produced from EMDB entry 5348). (**d**) Crystal structure of phage ϕX174; the capsid diameter is 33 nm (PDB entry 2BPA). (**e**) Electron micrograph of the filamentous phage M13; the filament is 8 nm wide (Reproduced from [7]. With permission). (**f**) Cryo-EM density of bacteriophage MS2 with a capsid of 28 nm in diameter (Reproduced from [8]. With permission)

measure protein, around which the tail is assembled and which thus determines the length of the tail.

A family of intermediate size DNA phages are the *Tectiviridae*. These bacterial viruses have the peculiarity of an internal membrane located below the outer protein shell and can be divided in two subgroups – those infecting Gram-positive and those infecting Gram-negative bacteria. Bacteriophage PRD1 is the most well known example [9]. It is a non-tailed icosahedral virus with a double-stranded DNA genome of 15 kilobases. PRD1 has a broad host range and infects several species of Gram-negative bacteria including *Escherichia coli* and *Salmonella*. Susceptibility of bacteria for PRD1 depends on the presence in the bacteria of a conjugative IncP, IncN or IncW plasmid, which codes for the receptor [10].

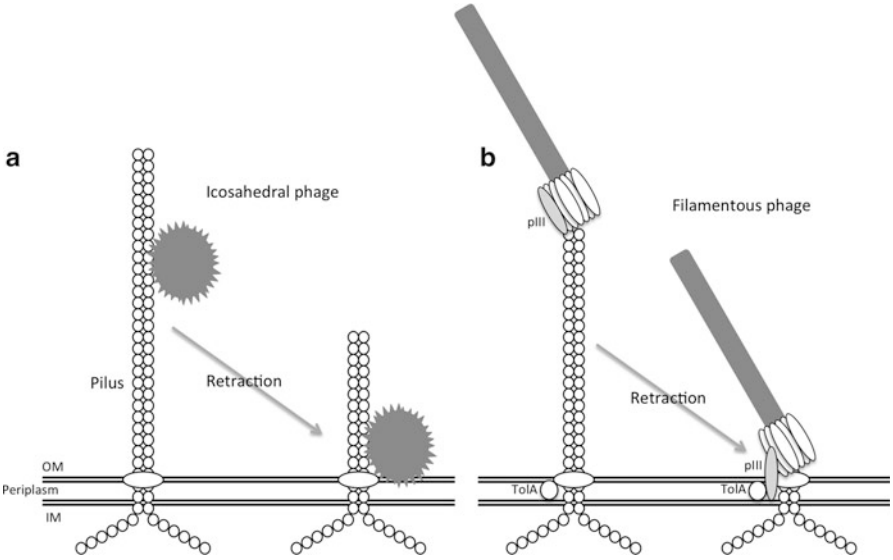


Fig. 17.2 Receptor binding of pilus-binding bacteriophages. (**a**) Icosahedral RNA phage (such as MS2 or Qβ) bound to the side of the pilus. After retraction of the pilus, the phage is close to the membrane and transfer of the RNA into the host takes place. (**b**) Filamentous RNA phage (such as M13 or fd) bound to the end of the pilus. After retraction of the pilus, pIII interacts with the secondary receptor TolA and transfer of the DNA into the host takes place

Small DNA phages or isometric phages have a covalently closed circular plus-strand DNA of up to 10 kilobases. They have an icosahedral capsid of around 30 nm. The prototype is ϕX174 (Fig. 17.1d), for which the different steps of its replication cycle (infection, DNA replication, capsid assembly and host lysis) have been studied [11, 12].

Filamentous phages also contain a circular single-stranded plus-strand DNA molecule that is folded back onto itself in the elongated phage particle. Prototypes of this family of phages are fd and M13 (Fig. 17.1e). They are useful cloning vectors for producing single-stranded DNA due to the lack of a packaging limit – larger circular DNA molecules simply lead to longer phage particles.

RNA phages are small, have icosahedral capsids and contain a single linear single-stranded RNA of about 4 kilobases, coding for only four genes (maturation protein, capsid protein, lysis protein and replicase protein). A well-studied RNA phage is MS2 (Fig. 17.1f) [13]. Like the filamentous phages, they attach to the F-coded pilus (Fig. 17.2) and can thus only infect bacteria containing the F-plasmid. However, they enter bacterial cells by attachment to the side of the F-pilus (Fig. 17.2a), rather than the end as filamentous phages do (Fig. 17.2b).

Phages may change their host range by mutation of their primary receptor-binding proteins or adoption of new receptor-binding proteins *via* horizontal gene transfer, and bacterial mutants that are resistant to infection by a particular bacteriophage have generally lost or modified the primary phage receptor. In this chapter, we will discuss

the specialised proteins used by bacteriophages for the very first step of infection, receptor recognition. We mention what is known about the mechanism different bacteriophages use to decide a correct host cell has been encountered and the phage should commit itself and proceed to the following step, DNA transfer. We will also revise what is known about the DNA transfer step for the different phages.

17.2 Proteins Used in Receptor Recognition and Nucleic Acid Transfer

Phages take advantage of host physiology for their infection process and replication. Like eukaryotic viruses, some bacteriophages recognise their host cell receptors with proteins of the capsid itself. Others have specialised proteins for this function. In this section, we discuss the receptor-binding mechanism of one or several examples of each type of phage. We will also discuss nucleic acid transfer for the different types of phages, although for many phages the mechanism of nucleic acid transfer is still unknown. Naming of bacteriophage proteins may be confusing to readers unfamiliar with phages. In many cases gp (gene product) followed by letters (capital or not) or numbers (arabic or roman) is used, *i.e.* gp37 for phage T4 or gpF in case of ϕX174. The single letter p or P (protein) may also be used, *i.e.* pIII of bacteriophage M13. In the case of bacteriophage T5, pb (protein band) is used, *i.e.* pb1. Proteins may be numbered or lettered according to the order of discovery, size, location on the genome or other scheme the discoverers fancied. Other names or the designation ORF (open reading frame) followed by a number may also be used.

17.2.1 RNA Phages

RNA phages contain 90 dimers of capsid protein, together forming a T = 3 icosahedral capsid [13]. They bind to the bacterial pilus, which is made up of multiple copies of the pilin protein, in the first step of infection. For MS2 (Fig. 17.1f), it is known that one copy of the pilin-binding maturation protein is located at one of the fivefold vertices. Cryo-EM reconstruction at 2 nm resolution on pilus-bound MS2 phage showed bound phages are attached at a slight angle, consistent with the symmetry-mismatch of a single copy of the maturation protein bound to one of the fivefold vertices [14]. Information at the atomic level on how the maturation protein interacts with pilin is not available. Retraction of the pilus presumably brings the phage to the bacterial membrane (Fig. 17.2). It is not known if specific secondary receptor interaction of the phage with membrane components takes place.

Apart from binding pilin on the outside of the capsid, the maturation protein is thought to bind the 5′ and 3′ ends of the viral RNA on the inside. Arrival of the phage at the membrane may be a signal for unpacking of the RNA and transferring it into the cell, but details of this process are unknown. At the vertex contacting the pilus, density that presumably includes contributions from both RNA and maturation protein is poised near the channel that goes through the fivefold vertex to the outside. This suggests that the RNA-maturation protein complex leaves the capsid as the first step of the nucleic acid transfer process [14], presumably through the hole observed at the fivefold vertex.

17.2.2 Filamentous Phages

Filamentous bacteriophages (Fig. 17.1e) are thin and long particles (about 8 by 900 nm) that are assembled around a circular single stranded DNA, they can be significantly longer if larger DNA molecules are encapsulated. The end of the virus that assembles first is formed by proteins pVII and pIX, this end is also called the distal end. The major capsid protein is protein pVIII, a small α-helical protein that assembles in helical fashion around the encapsulated DNA molecule. The proximal end is formed by proteins pVI and pIII. The receptor-binding protein pIII has three domains, N1, N2 and a C-terminal domain that anchors the protein to the phage. Infection initiates by absorption of protein pIII to the distal terminus of F-pilus (Fig. 17.2b), *via* the N2 domain of pIII. As both the tip and sides of the conjugative pilus are formed by the protein pilin, filamentous phages must recognise a different epitope than the small RNA phages – an epitope that is not available on the sides of the pilus.

The C-terminal domain of the *Escherichia coli* periplasmic protein TolA acts as co-receptor for filamentous phages [15]. Upon retraction of the pilus by the bacterium, the phage protein pIII approaches the membrane and a normally shielded epitope of domain N1 of pIII binds to TolA (Fig. 17.2b). Details of subsequent steps are obscure, although it has been postulated that pIII may participate in forming a channel through the bacterial membrane [15]. It is not known whether only naked phage DNA enters the bacterium or whether phage capsid proteins also enter.

17.2.3 Small DNA Phages

Small DNA phages, also called isometric phages, recognise their receptor by the capsid proper. In the case of ϕX174 (Figs. 17.1d and 17.3), primary, reversible receptor interaction with the terminal N-acetyl-glucosamine at the non-reducing end of the core polysaccharide of bacterial lipo-polysaccharide (LPS; [11]) is mediated by the capsid protein gpF. The binding site has been localised to the

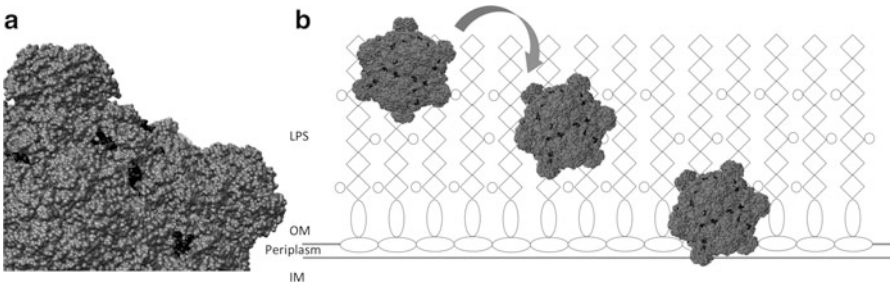


Fig. 17.3 Receptor binding of bacteriophage ϕX174. (**a**) Location of the primary, reversible, receptor binding site (in *black*) on the phage capsid (*grey*) formed by protein gpF. (**b**) Steps in attachment and DNA transfer. Phage ϕX174 binds reversibly to the non-reducing ends (*grey circles*) of the core polysaccharide of bacterial lipo-polysaccharide. ϕX174 may roll over the cell surface until a suitable, unknown, secondary receptor is encountered. After irreversible binding to this receptor, transfer of the pilot protein gpH bound to phage DNA into the host takes place

side of a depression on gpF [16]. Spike proteins gpG and gpH are necessary for irreversible cell attachment to a second factor [12], which may be encoded by the *phxB* gene of *Escherichia coli*. Usually, two to three spikes become embedded in the cell wall and the phage submerges to about one-half of its diameter (Fig. 17.3). Tailed bacteriophages are thought to diffuse laterally along the cell surface after reversible attachment until they encounter their secondary receptor and bind irreversibly. Bacteriophage ϕX174 cell attachment may proceed similarly. However, instead of walking along the surface, ϕX174 may roll (Fig. 17.3b).

Adsorbed ϕX174 particles are often located at regions of adhesion between the cell wall and the inner membrane. This suggests that the secondary receptor is also located here and indicates that the DNA genome may be transferred directly into the cytoplasm. DNA eclipse from the phage has a high activation energy barrier and requires the LPS lipid A moiety of the host. DNA transfer into the cell requires protein gpH of the phage (also called the DNA pilot protein) and an active host cell metabolism [11]. This protein also enters into the cell and directs stage I DNA synthesis. It has been proposed that DNA synthesis is necessary for complete DNA transfer, in the same way that active transcription is necessary for transfer of phage T7 DNA (see below). In the atomic structure of ϕX174, diffuse density is located at each of the fivefold vertices of the virion. This density has been attributed to the DNA pilot protein [16], suggesting that DNA exits through one of the fivefold vertices. After entering into the bacterium with the DNA, GpH promotes phage replication, perhaps by stimulating transcription, being involved in DNA replication or just stabilising the DNA and protecting it against degradation by host cell enzymes.

17.2.4 Lipid-Containing Icosahedral Phages

The best studied and prototype of the *Tectiviridae*, the virulent bacteriophage PRD1 (Fig. 17.4f), first reversibly binds to its receptor, followed by an irreversible step

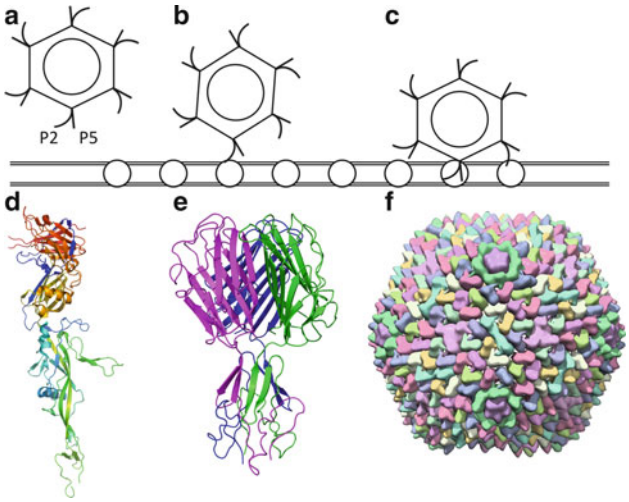


Fig. 17.4 Receptor binding of bacteriophage PRD1. Schematic drawings of (**a**) phage PRD1 and its spike proteins P2 and P5, (**b**) phage PRD1 binding to its receptor using the P2 protein and (**c**) opening of the vertex for membrane penetration and DNA ejection. (**d**) Crystal structure of the monomeric receptor-binding protein P2 (PDB entry 1n7v). (**e**) Crystal structure of the carboxy-terminal distal part of the trimeric vertex protein P5 (PDB entry 1yq8). (**f**) Structure of bacteriophage PRD1 (based on PDB entry 1W8X)

(Fig. 17.4a–c). The vertex of bacteriophage PRD1 shows two different types of spikes associated with the pentameric P31 capsid protein, composed of the proteins P2 (Fig. 17.4d) and P5 (Fig. 17.4e) [10]. Specific binding to the cell surface receptor depends on the protein P2. The cell surface receptor is a component of the conjugative system coded for by a plasmid, as bacteria without these plasmids are resistant to PRD1 infection. The entire conjugative apparatus coded by the plasmid is necessary for infection, and the TraF protein is considered to be the phage receptor protein. Receptor binding of the P5 protein has not been demonstrated, but it might give the virus a selective advantage in the wild by binding to an unknown, alternative receptor.

After binding to the receptor, P2 is thought to trigger conformational changes within the spike complex that are transmitted to the internal membrane. Information about receptor binding is somehow transferred to the DNA delivery apparatus, which contains at least the proteins P7, P11, P14, P18 and P32. This leads to considerable conformational changes in the vertex structure. The removal of the spike complex creates an opening in the vertex, which enables an appendage to protrude that penetrates the outer membrane of the host. The protruding tube probably contains protein P11, but may also contain an extension of the phage lipid membrane, as isolated phage membrane forms small vesicles with tubular extensions [9]. The lytic transglycosylase (P7) is thought to assist in genome entry by locally degrading the peptidoglycan layer. The appendage formed extends some 35 nm (the thickness of the cell envelope) penetrating the cytoplasmic membrane.

DNA transfer is dependent on active membrane tube formation and reduction of the membrane vesicle volume assisted by at least the proteins P14, P18 and P32. The force required for DNA transfer may result in part from the pressure release of the tightly packed DNA in the capsid, but it is tempting to speculate that contraction of the lipid membrane of the phage also plays a role. After all, in the phage the membrane is forced into a spherical form, while in isolation it prefers to assume a tubular conformation. After DNA injection, transcription and protein-primed genome replication take place. It is not known whether, like in the case of bacteriophage T7 (see below), transcription may also play a role in DNA transfer. The phage protein shell stays associated with the host cell surface.

17.2.5 Large DNA Phages

Large double-stranded DNA phages of the *Caudovirales* order possess a tail for efficient host cell recognition and DNA transfer (Fig. 17.1a, b, and c). In general, in a first step reversible attachment is achieved, with a subsequent irreversible attachment, after which the phage is fully committed to DNA transfer into the selected host cell. Proteins on the outside of the DNA-containing capsid may also be involved in highly reversible attachment to the host, augmenting the local phage concentration near the bacterium and favouring efficient infection. In the following three sections, we will discuss what is known about receptor binding and DNA transfer for the most known examples of the *Podoviridae* (P22 (Fig. 17.1c), T7 and ϕ29), *Siphoviridae* (T5, λ (Fig. 17.1b), SPP1 and lactobacillus phages) and *Myoviridae* (T4 (Fig. 17.1a), P2, SPO1 and spore-binding phage 8a).

Podoviridae

Podoviruses have a tail that is not long enough to span the periplasmic space of Gram-negative bacteria or the cell wall of Gram-positive bacteria. Here we will discuss the host recognition and DNA transfer steps of three podoviruses as examples, two that infect the Gram-negative bacteria *Salmonella enterica* (phage P22) and *Escherichia coli* (phage T7), and one that infects the Gram-positive bacterium *Bacillus subtilis* (phage ϕ29).

Bacteriophage P22. P22 (Fig. 17.1c) infects the host cell using a three-step mechanism (Fig. 17.5a–c): (i) the virus binds through its tailspikes to LPS; (ii) the tail needle contacts with the secondary receptor and/or pierces the membrane; (ii) the tail proteins change their conformation and the DNA is ejected. As in many other podoviruses, the P22 tail (Fig. 17.5d) contains six tailspikes (Fig. 17.5e). Each one is a stable homotrimer of gp9 [17], and can be divided into three parts: an amino-terminal phage-binding domain, a central parallel β-helix domain and a carboxy-terminal, highly interdigitated part important for trimerisation and thermostability. The amino-terminal phage-binding domain is more evolutionary conserved between related phages, while the sequence of the carboxy-terminal domains has evolved more, apparently depending on the bacteria to infect. In phage P22 a hinge is present between the amino-terminal domain and central

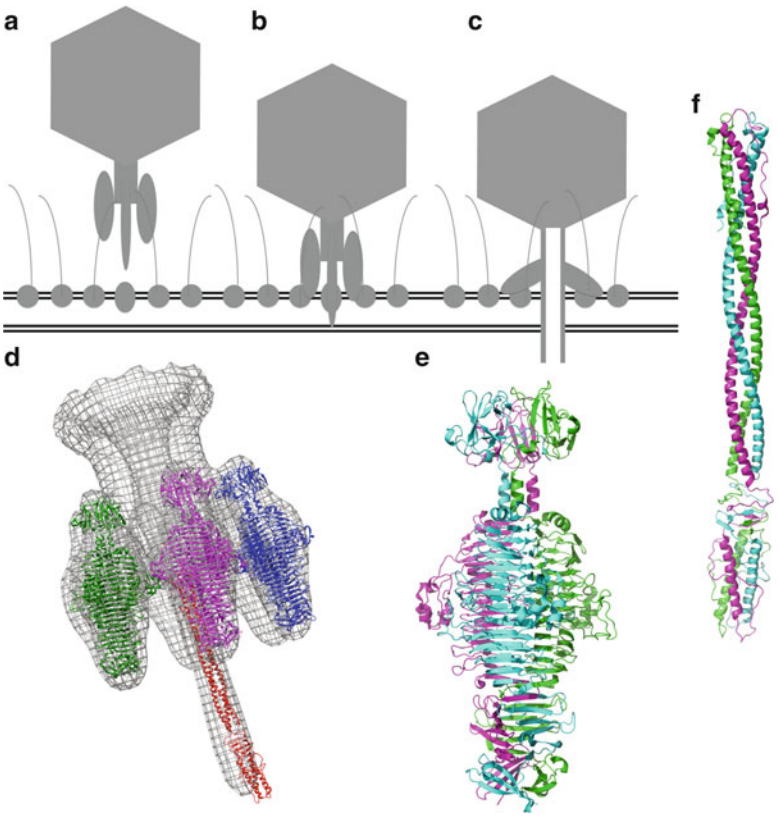


Fig. 17.5 Bacteriophage P22. (**a**) The bacteriophage binds through its tailspikes to LPS. (**b**) The tail needle contacts with the secondary receptor and/or pierces the membrane. (**c**) The tail proteins change their conformation and the DNA is ejected. (**d**) Cryo-EM reconstruction (EMDB: 1119) of the tail with crystal structures of tailspikes and tail needle fitted. (**e**) Crystal structure of the tailspike (gp9, PDB: 2XC1). (**f**) Crystal structure of the tail needle (gp26, PDB: 2POH)

domain, which allows flexibility between the two. Folding and trimeric assembly of the tailspikes are dependent on their interdigitated carboxy-terminal domains. The tailspike possesses receptor-destroying endorhamnosidase enzymatic activity. It cleaves the α(1,3)-O-glycosidic bond between rhamnose and galactose of the bacterial LPS O-antigen repeats. This results in cleavage of the LPS O-antigen receptor during the adsorption of the phage to the cell surface, rendering the binding reversible. More importantly, this activity may be used by the phage to clear a path through the bacterial outer LPS region (the capsule) to gain access to the secondary receptor on the cell surface. The enzymatic activity may also serve to release progeny phage particles from the cellular debris at the end of the lytic cycle.

From the centre of the tail, the needle protein projects, which is a thin trimer of gp26 [18] (Fig. 17.5f). The tip of the needle projects beyond the tailspikes (Fig. 17.5d) and is thus likely the first protein to contact the cell surface. Whether

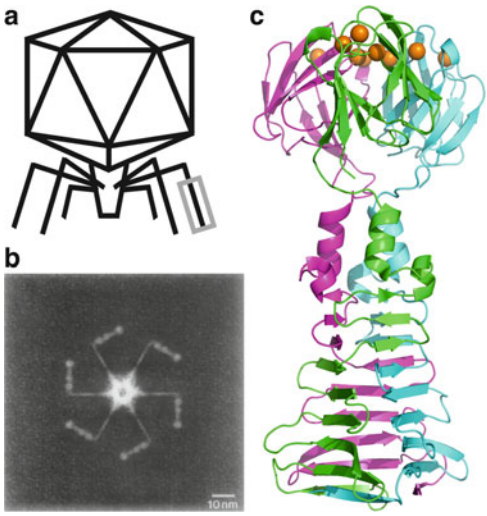


Fig. 17.6 Bacteriophage T7. (**a**) Schematic drawing of phage T7 (Reproduced from [21]. With permission). (**b**) Bottom view of the tail, which clearly shows the six fibres (Reproduced from [22]. With permission). (**c**) Crystal structure of the carboxy-terminal domain of the fibre (PDB entry 4A0T) with residues that affect host range marked with orange balls

this interaction is specific, *i.e.* with a putative secondary receptor, or unspecific, for example with the membrane, is currently unknown. At the same time, gp26 is the plug that keeps the DNA in the P22 particle; which suggests that this interaction is somehow connected to DNA release. Bacteriophage P22 contains three so-called ejection proteins: gp7, gp16 and gp20 [19]. These proteins are likely involved in DNA transfer from the phage into the host bacterium. However, the mechanism is still unknown.

Another *Salmonella* phage that has been studied quite extensively is ε15. Electron cryotomographic analysis of the infection process of *Salmonella anatum* by ε15 has shown that in a first stage, the tailspikes of ε15 attach to LPS on the surface of the bacterium. The tail hub of ε15 then attaches to an unknown cell receptor and forms a tube. Transferred core proteins and cellular proteins may contribute to the tube formation, which spans the periplasmic space. DNA transfer then takes place through the tube. The tube allows the DNA to directly enter the cytoplasm and protect it from periplasmic nucleases. Once the DNA has been injected into the cell, the tube and portal seals, and the empty bacteriophage remains at the cell surface [20].

Bacteriophage T7. Not all podoviruses contain enzymatically active receptor binding tailspikes. Instead, T7 contains six thin, kinked, tail fibres (Fig. 17.6). The fibres, homotrimers of the protein gp17, are responsible for the first specific, but still reversible, attachment to *Escherichia coli* LPS. The kinked fibres are comprised of an amino-terminal tail attachment domain, a slender shaft, a flexible kink, and a carboxy-terminal domain composed of several nodules [22]. The structure of the carboxy-terminal region comprising residues 371–554 is known [23]. The structure revealed a β-stranded pyramid domain and a globular carboxy-terminal receptor

binding domain, in which residues known from mutational studies to be involved in receptor binding are located in loops at the end of the trimer (Fig. 17.6c).

Binding of the T7 fibre to the inner core of the LPS mediates the initial contact with the target cell, but does not yet trigger the DNA release. This initial and reversible binding may position the phage for irreversible binding and DNA ejection. Whether a specific receptor is involved in the secondary, irreversible, binding, is currently unknown. If so, the first, reversible binding may allow the phage to stay associated with the bacterial surface while it diffuses two-dimensionally until the putative secondary receptor is encountered. For T7, these secondary interactions may be mediated by its outer tail proteins gp11 and/or gp12. Penetration of the outer membrane of a Gram-negative bacterium by a podovirus must involve forming a channel for DNA transport across the cytoplasmatic membrane [19]. The T7 tail is too short to span the periplasmic space of *Escherichia coli* and a channel needs to be made to allow the phage genome to travel from the virion to the cytoplasm. It is thought T7 extends its tail by ejecting the internal head proteins (gp14, gp15, gp16) into the cell prior to DNA transfer into cytoplasm. These three proteins are known to form a cylindrical structure inside the phage head, and they are ejected from the infected virion into the cell envelope before the phage genome. They should disaggregate from their structure in the mature virion and also almost completely unfold in order to leave the head and pass through the head-tail connector and form a trans-envelope channel that connects the virion tail tip to the cell cytoplasm. Bacteriophage T7 has therefore been described as having an extensible tail.

After phage adsorption and protein ejection, approximately one kilobase of the 40-kilobase genome is normally internalised by the cell [24], perhaps as a result of the release of the pressure of the DNA encapsulated in the capsid (see Chaps. 12, 18 and 19). Subsequent to internalisation of the leading first kilobase, translocation is coupled to transcription. The leading three kilobases include mainly promoters for *Escherichia coli* RNA polymerase. The RNA polymerase moves along T7 DNA until it encounters the cytoplasmatic membrane but continues to transcribe, thereby pulling DNA from the capsid into the cell. Transcription by *Escherichia coli* RNA polymerase produces messenger RNA for the T7 proteins located at the leading end of the genome, including the single-subunit T7 RNA polymerase and an inhibitor of bacterial restriction enzymes that might otherwise degrade the T7 DNA. The leading end of the genome does not contain restriction sites, while the rest of the genome does. At 30 °C, the T7 genome is internalised at a constant forty base pairs per second across the genome, the same rate as transcription of ribosomal RNA operons. *Escherichia coli* RNA polymerase is one of the strongest motor proteins known and it is thus not surprising that it can pull a DNA molecule into the cell cytoplasm. Of course, once T7 RNA polymerase is produced, it may well contribute to DNA internalisation, although it is not absolutely necessary for this function.

Many encapsulated bacteria are infected by specialised bacteriophages that carry host-capsule degrading hydrolases. These usually tailspike-associated enzymatic activities enable the phages to penetrate the capsular layer and are important determinants of the bacteriophage host range. Podovirus K1F infects *Escherichia coli* K1, a polysialic acid encapsulated pathogen. K1F has degrading tailspikes known as endosialidases that break the α2,8-glycosidically linked sialic acid

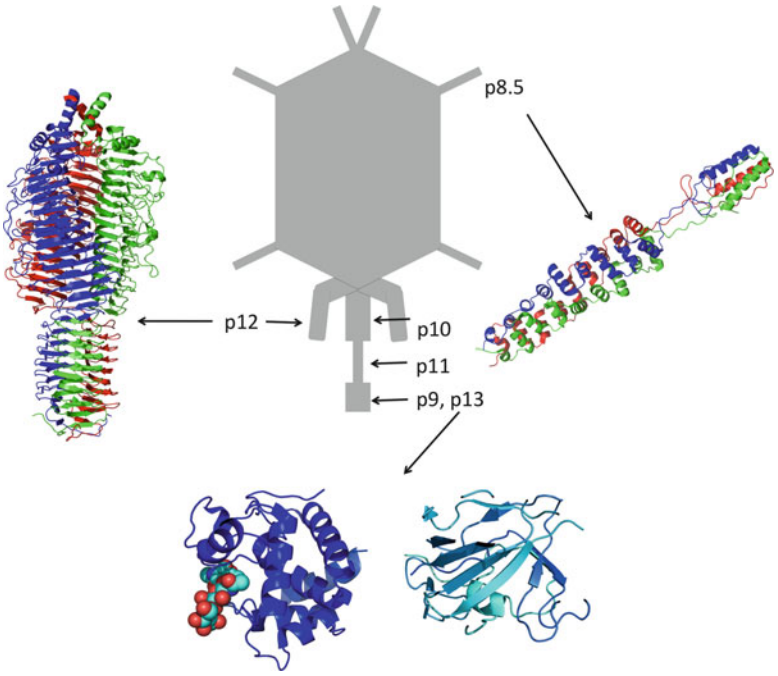


Fig. 17.7 Bacteriophage ϕ29 and its constituent proteins important in the first steps of infection. A schematic drawing of phage ϕ29 is presented with p8.5, p10, p12, p11, p9 and p13 indicated. The phage is surrounded with crystal structures of the relevant proteins. Crystal structures are shown of the head fibre p8.5 (PDB entry 3QC7), the carboxy-terminal part of the appendage p12 (PDB entry 3GQ7) and p13. The amino-terminal lysozyme-like domain and the carboxy-terminal endopeptidase-like domain of p13 are shown (*left* and *right*, respectively), the amino-terminal domain with the co-crystallised substrate tri-N-acetyl-D-glucosamine (PDB entries 3CSZ and 3CSQ, respectively). P13 is located at the end of the phage tail

oligomers or polymers. The tailspike endoNF permits the phage to adsorb to the polysialic capsule, depolymerise the capsule to tunnel towards the outer bacterial membrane and to mediate the adhesion to the membrane. Infection may then proceed as for other podoviruses. There are only some isolated phages capable of degrading biofilms. The T7-related bacteriophage ϕ15 is able to degrade biofilms of two *Pseudomonas putida* strains indicating exopolysaccharide depolymerase activity. The ϕ15 tailspike protein is responsible for exopolysaccharide degrading activity and this exopolysaccharide is probably a primary and essential receptor for ϕ15.

Bacteriophage ϕ29. As bacteriophage ϕ29 infects a Gram-positive host, it is perhaps not surprising it has some specific characteristics compared to phages infecting Gram-negative hosts. The phage capsid is decorated with 55 head fibres attached to the quasi-threefold symmetry positions (Fig. 17.7). The head fibres are trimers of the protein p8.5. Although phage particles without head fibres are infective, the head fibres might enhance attachment of the virions onto the host cell wall, interacting with the cell teichoic acids [25].

The tail of bacteriophage ϕ29 consists of the dodecameric portal protein p10, 12 radially arranged trimeric appendages formed by p12 and an extension formed by proteins p11, p9 and p13 (Fig. 17.7). Protein p12 can cleave the teichoic acid of the bacterial cell wall and is thus the functional equivalent of the P22 tailspike in ϕ29 in mediating the first specific, but reversible binding of the phage to the bacterial host. The p12 appendages are kinked; an amino-terminal cylindrical outward-pointing arm links them to the lower collar region of the ϕ29 tail and a carboxy-terminal spindle-shaped domain contains the polyglycerol phosphate degrading active site [26]. Most of the carboxy-terminal enzymatic domains point downwards in a cryo-EM reconstructed structure of the whole phage, but two point horizontally outwards. The movement of the appendages may allow the phage to move downwards through the teichoic acid layer to the bacterial membrane [27].

Protein p12 monomers are β-helical and assemble into a homotrimer in a sideways manner (Fig. 17.7). An intramolecular chaperone at the very carboxy-terminus helps the p12 monomers to trimerise and is then cleaved off by an autoproteolytic reaction to yield the mature protein. Proteins p11, p9 and p13 make up the central rod of the tail, with p13 being the most distal protein and thus the one likely to make the first irreversible contact with the bacterial cell surface. Its crystal structure revealed both a transglycolase domain and a metalloendoprotease domain. Both likely function in degrading the peptidoglycan layer before DNA transfer can occur [28].

Bacteriophage ϕ29 DNA transfer proceeds *via* a push-pull mechanism ([29] and references therein). The right end of the genome enters the bacterium first and the leading two-thirds enter the cell in the absence of phage protein synthesis, presumably by release of the pressure of the packaged DNA in the capsid (the push phase). Transfer of the remaining third of the genome depends on synthesis of the specific phage proteins p16.7 and p17 encoded by the part of the genome transferred in the push phase. Energy is also required for this pull phase, but the exact mechanism is unknown.

Siphoviridae

Siphoviruses have an apparent communication problem between the two ends of their long, flexible tail tube. Somehow, the fact that a suitable bacterium to infect has been encountered with the distal end of the tail has to be transmitted to the end most proximal to the head domain, where the packaged DNA must be released. A conformational change taking place in a domino effect, opening the inner diameter of the tail, has been proposed for this purpose [30], although a simpler mechanism may be that opening the bottom of the tail allows the tape measure protein to leave and DNA simply follows [31].

Bacteriophage T5. Like many phages that infect Gram-negative bacteria, bacteriophage T5 (Fig. 17.8a) has two different types of fibres: the outward-pointing L-shaped tail fibres and the central straight tail fibre [32]. The three L-shaped tail fibres are trimers of the protein pb1. They bind reversibly to the O-antigen of *Escherichia coli* LPS and accelerate the absorption [33]. However, loss of this first receptor does not lead to T5 resistance. The interaction of the L-shaped tail fibres with O-antigen accelerates the adsorption of the phage to the cell and keeps the fibred phage at the surface of the host for a considerably longer time than T5 phage

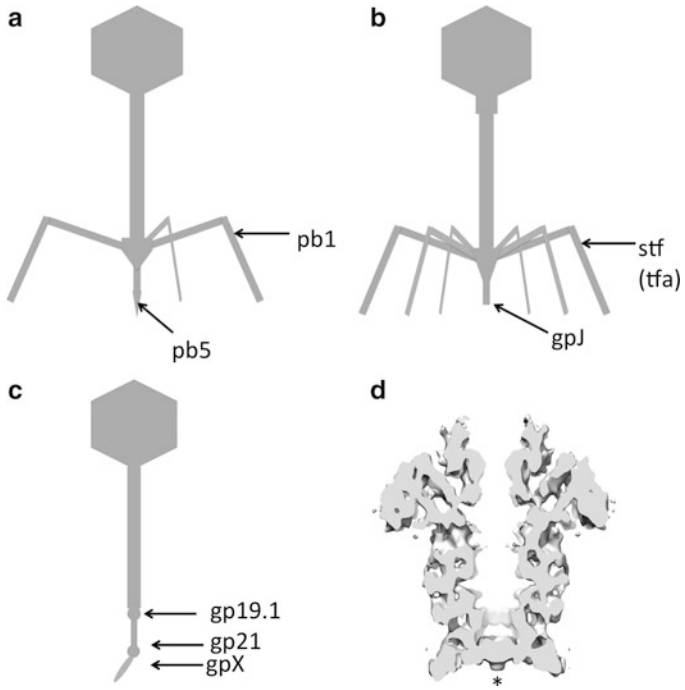


Fig. 17.8 Schematic drawings of some siphoviruses. (**a**) Bacteriophage T5 with the L-shaped tail fibres (pb1) and the FhuA-binding protein pb5 indicated. (**b**) Bacteriophage λ with the side tail fibres (stf, tfa) and the LamB-binding protein gpJ indicated. (**c**) Bacteriophage SPP1. The positions of the hub protein gp19-1, the tail protein gp21, which may have a glycan hydrolysing function, and the unknown receptor-binding tail tip, gpX, are indicated. (**d**) Cut-through of the neck region of phage SPP1 consisting of the portal protein gp6, gp15 and gp16. Electron microscopy density (EMDB 1021) is shown in *grey*. An *asterisk* indicates the stopper of gp16 that prevents DNA from exiting the phage head before infection

lacking fibres, making binding to the secondary receptor more likely the longer the phage can move along the cell surface. Efficient surface attachment occurs when more than one of the three tail fibres binds to an O-antigen at the same time [34]. The L-shaped tail fibres have a thin outward-pointing part followed by a kink and a thicker downward-pointing part consisting of consecutive bead-shaped domains. Pb1 is a large protein of 1,396 amino acids per monomer. The first 200 amino acids are rich in alanines and serines and are predicted to form a coiled-coil, which may correspond to the thin domain of the L-shaped tail fibres. The remaining part is more varied in sequence and contains large hydrophobic and aromatic residues, so likely forms the consecutive globular domains. The very carboxy-terminal part of the protein is an intramolecular chaperone and probably functions in correct trimerisation of the protein as it does for p12 of phage ϕ29, the carboxy-terminal 132 amino acids probably get autoproteolysed after trimerisation. No atomic structure for any of the domains of pb1 is known, but like other phage fibres it is likely to contain β-structure.

In a second adhesion step, pb5 binds irreversibly to the outer membrane ferrichrome transporter FhuA, a monomeric transporter for which the crystal structure is known. The gating loop of FhuA, consisting of residues 316–356, is the largest extracellular loop and is crucial for the binding of the ferrichrome-iron and T5 phages. After FhuA binding, the straight fibre tip of the viral tail traverses the lipid bilayer and undergoes a major conformational change [35]. A pore is formed in the bacterial membrane next to FhuA through which the DNA is transferred. In the mature phage particle the DNA is connected to the proximal end of the tail. Upon binding to the receptor a signal has to be transmitted from the receptor-binding protein along the tail to the head-tail connection to induce DNA penetration through the tail tube. Signal transmission may occur as a progressive conformational change of the major tail protein starting at the distal end of the tail tube, perhaps widening the inner volume to allow for DNA passage. Bacteriophage T5 transfers its double-stranded DNA in two steps. In a first step, 8 % of the total T5 chromosome is transferred, perhaps by simple pressure release of the tightly packed DNA in the capsid. After a pause of a few minutes during which proteins coded by this fragment are synthesised, the remaining part of the DNA is transferred [36]. The section of the T5 genome that is injected in the first step encodes proteins that degrade host cell DNA and shut down the synthesis of host proteins. How the remaining section of DNA is transferred into the bacterium is unknown.

Bacteriophage λ. The lysogenic bacteriophage λ (Figs. 17.1b and 17.8b) has traditionally been an important model system for molecular biology. Laboratory strains of λ appear to have lost their side tail fibres and can still infect their host *Escherichia coli* efficiently in laboratory conditions [37]. However, λ Ur (the original λ) does have side tail fibres (stf; Fig. 17.8b). The side tail fibres need the tail fibre assembly protein tfa for their correct trimerisation and assembly; it is not sure whether tfa remains associated with the side tail fibres in the mature phage particles. The side tail fibres probably recognise OmpC in a primary, reversible binding step [38]. In a second step, the carboxy-terminal end of gpJ at the tip of the tail binds to the *Escherichia coli* lamB gene product [39], a maltoporin. These two proteins form a very stable complex and this interaction probably leads to irreversible binding. Two types of complexes have been identified of the λ tail tip with LamB-containing membranes. In the first, the hollow tail remains at a distance of 17 nm from the membrane. In a subsequent step, the hollow tail attaches to the membrane for DNA transfer.

During DNA transfer, the linear phage genome is transferred past the *Escherichia coli* outer membrane. During this process, transmembrane channels are formed, which permit the entry and escape of small molecules, but not proteins [40]. The DNA then passes through a separate bacterial sugar transport protein, the mannose permease of the bacterial phosphotransferase system, in the cytoplasmic membrane [41]. The two hydrophobic subunits II-PMan and II-MMan alone are sufficient for penetration of λ DNA. It seems thus that bacteriophage λ has subverted the bacterial maltose entry system into a transfer system for its DNA. Once in the cytoplasm, the linear DNA circularises using the *cos* sites, which are the twelve-base G-C rich cohesive sticky ends of the genome.

Bacteriophage SPP1. The Gram-positive *Bacillus subtilis* phage SPP1 (Fig. 17.8c) has a single central fibre for host interaction. The tail tip of SPP1 binds to YueB, a membrane protein whose extracellular region crosses the thick cell wall to expose a receptor region at the bacterial surface. The tail tip also mediates degradation of the bacterial cell wall. The tip is 31 nm long and 10 nm wide at its largest diameter. It can be divided into three domains: a sphere-like region, a broad flattened domain and a terminal rod (Fig. 17.8c). These domains correspond to different proteins: gp19.1 being the sphere-like region nearest the phage, gp21 forming the central part and an as yet unknown distal receptor-binding protein. Gp19.1 is the baseplate hub protein and forms a hexameric ring with a central hole [42]. Gp21 has predicted β-helical structure and may resemble the P22 tailspike in structure and glycan hydrolysing function.

The tip does not have a DNA channel and therefore probably moves out of the way before DNA transfer. The binding of the tail fibre to YueB leads to the irreversible commitment of the virus particle to eject its DNA. In this case, no previous, reversible, interactions have been identified. The structure of the siphovirus SPP1 tail has been determined before and after DNA ejection [30]. The results revealed extensive structural rearrangements in the internal wall of the tail tube. A proposal was made that the binding of the tail tip to YueB triggers a conformational switch that is propagated as a cascade of conformational changes along the 160 nm-long helical tail structure to reach the head-to-tail connector. This would lead to the connector opening and allowing passage of DNA into the host cell through the tail tube. The structure of the connector region between the tail and the head of DNA-containing phage has been determined by electron microscopy and fitting of atomic structures. It consists of three dodecameric proteins: the portal protein gp6, the adaptor protein gp15 and a "stopper" protein, which prevents premature DNA release [43] (Fig. 17.8d).

***Lactococcus lactis* bacteriophages.** Phages of the Gram-positive *Lactococcus lactis* bacterium sport more compact receptor-binding proteins. The crystal and electron microscopy structures of the baseplates of two *Lactococcus lactis* phages have been determined (Fig. 17.9): that of lactococcal phage p2 (not to be confused with the Gram-positive bacteria-infecting myovirus P2) and TP901-1. The baseplate of phage p2 is composed of three protein species [44]. The central part of the baseplate is formed by a circular hexamer of ORF15 with a central hole. A trimer of ORF16 is located at the bottom of the baseplate, forming a closed dome that does not allow DNA passage. Six ORF18 trimers are attached to the central ring, each trimer interacting with a carboxy-terminal extension of an ORF16 monomer. ORF18 is the receptor binding protein and its structure shows an N-terminal β-sandwich shoulder domain, which binds to the baseplate. The remainder of the structure is a short triple β-helical neck and a carboxy-terminal head domain. In the unactivated baseplate, the six trimeric receptor-binding proteins point upwards, *i.e.* away from the bacterial membrane. Activation by calcium ions causes large conformational changes in p2 baseplate, leading to rotation of receptor-binding domains by 200° to point downwards towards the host cell. The head domains of ORF18 may recognise lipoteichoic acids, which are phospho-glycerol polymers. The conformational change of the baseplate also leads the three ORF16 monomers

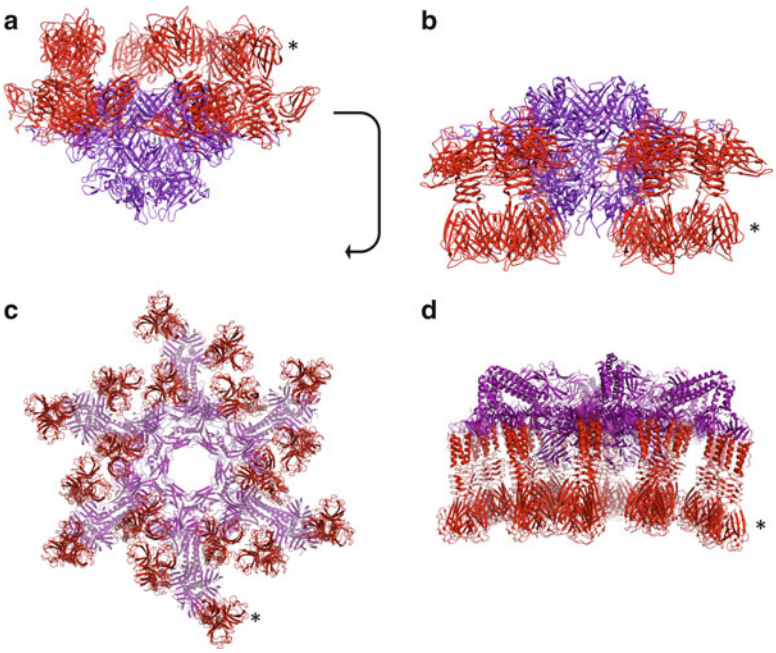


Fig. 17.9 Crystal structures of *Lactococcus lactis* bacteriophage baseplates. (**a**) The baseplate of phage p2 before activation, with the receptor-binding domains of the receptor-binding proteins (in *red*) pointing upwards, away from the bacterium (PDB entry 2WZP). The movement of 200° by the receptor-binding proteins that completely reverses their orientation is indicated by an *arrow*. (**b**) The baseplate of phage p2 after activation by calcium ions, with the receptor-binding domains of the receptor-binding proteins (in *red*) pointing downwards, towards from the bacterium (PDB entry 2X53). (**c**) The baseplate of phage TP901-1 (PDB entries 4DIV and 4DIW) viewed from below (*i.e.*, from the side of the bacterium) and viewed from the side (**d**). The 18 receptor-binding protein trimers are in *red*. One of the 18 receptor-binding domains in parts (**a**) and (**b**), as well as one of the 54 receptor-binding domains in parts (**c**) and (**d**), are indicated by *asterisks*

to separate, opening up a hole in the centre of the baseplate, presumably allowing passage of DNA for transfer into the host.

The baseplate of phage TP901-1 is composed of multiple copies of four different proteins [45]. The centre is a hexameric circular core formed by the Dit protein. The hole in this core is filled by a central tail fibre, Tal. From the core, six arms emanate, each arm being composed of a trimer of the BppU protein. The arms are α-helical up to the elbow. The rest of the arm points downwards and is an adaptor domain for the receptor-binding proteins. To each adaptor, three trimeric receptor binding proteins bind, leading to a total of 54 sites for the host cell envelope saccharides. The receptor-binding proteins point downwards, *i.e.* towards the host bacterium, ready for adhesion. No calcium ions are necessary for activation of TP901-1, so there are likely no conformational changes that have to take place before receptor binding. How receptor binding is related to DNA transfer in these phages is less clear, perhaps the strong binding with up to 54 receptor molecules pushes the

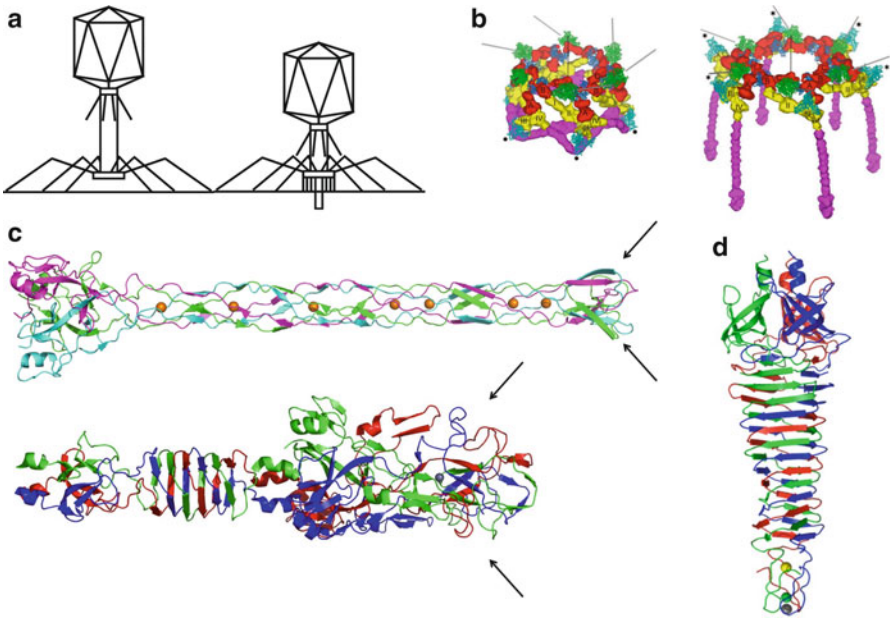


Fig. 17.10 Receptor recognition and DNA transfer by *Escherichia coli* myoviruses. (**a**) Schematic drawing of bacteriophage T4 before and after tail contraction (Adapted from [46], copyright by the National Academy of Sciences). (**b**) Structure of the baseplate in the hexagon and star conformations with the different proteins coloured differently. The short tail fibres are shown in *magenta*; folded away in the hexagon-state (extended tail sheath), and extended in the star-state (contracted tail sheath) (Reproduced from [47]). (**c**) Structures of the long tail fibre protein gp37 tip domain and the carboxy-terminal domain of gp12. *Arrows* indicate the receptor-binding regions. (**d**) The central spike of bacteriophage P2. The structure contains three ions at the tip: a calcium ion (*yellow*), a chloride ion (*green*) and an iron ion (*grey*)

central tail fibre against the cell wall and the force of the cell wall pushing against the tail fibre is sensed by the other end of the tail fibre, which opens a hole in the end of the tail tube.

Myoviridae

With their double-shafted tail, myoviruses (Fig. 17.1a) have resolved the host cell attachment and DNA transfer steps in perhaps the most advanced way, piercing the bacterial outer membrane and traversing the periplasmic space or peptidoglycan layer with the inner tail tube upon contraction of the outer tail tube (Fig. 17.10a). The tips of myovirus inner tail tubes even sport a spike for efficient membrane penetration [48] (Fig. 17.10d). This way, even large genomes of up to several hundred kilobases can be transferred efficiently into the bacterial cell. Here we discuss what is known about receptor attachment and DNA transfer of the most well known myoviruses.

Bacteriophage T4. Like the head fibres of phage ϕ29, immunoglobulin-like domains on the surface of bacteriophage T4 (Fig. 17.1a; see also Chap. 11) could interact with the cell surface and help to maintain the phage in the vicinity of the cell

surface by weak and non-specific interactions until the correct receptor is contacted. The bacteriophage T4 capsid is decorated with two non-essential outer capsid proteins: Hoc (highly antigenic outer capsid protein) and Soc (small outer capsid protein). Soc provides stability to T4 to survive under hostile environments; Hoc has immunoglobulin-like domains and may interact with the bacterial surface providing certain advantages [49].

The archetypal myovirus T4 (and presumably other myoviruses as well) infect their host bacteria efficiently compared with podovirus and siphoviruses, probably due to the sophistication of its contractile tail [50] (Fig. 17.10a). Infection of phage T4 starts with the interaction of the long tail fibres with cell surface receptors (see below). The long tail fibres of bacteriophage T4 are kinked structures of about 145 nm long with a variable width of up to 5 nm [47]. They can be divided into proximal and distal half-fibres, attached to each other at an angle of about 20°. Upon phage assembly in the previous host cell and in adverse conditions for phage multiplication, the long tail fibres are in a retracted conformation, lying against the tail sheath and head of the bacteriophage. Because the neck whisker protein (called gp*wac* or fibritin) is necessary for incorporation of the long tail fibres into the phage [51], we can suppose that this interaction is also responsible for maintaining the long tail fibres in a retracted state. In the extended conformation, only the proximal end of the fibre is attached to the baseplate and the fibre tips point outwards.

The long tail fibres are composed of four different proteins: gp34, gp35, gp36 and gp37 [47]. The proximal half-fibre (the thigh) is formed by a homotrimer of gp34. The amino-terminal part of gp34 is attached to the trimeric protein gp9 of the baseplate while the carboxy-terminal part interacts with the gp35. Gp35 (the knee) is a monomer and may be responsible for creating the angle between the proximal and distal half-fibres. The distal half-fibre is composed of a trimer of gp36 and a trimer of gp37. Gp36 forms the upper part of the shin and gp37 constitutes the rest of the shin, including the very distal receptor recognsing tip (the foot). Only the atomic structure of the receptor recognising tip is known [46]. It forms a globular collar domain, followed by a thin needle region in which seven iron atoms are coordinated and ends in a small intertwined receptor-binding domain (Fig. 17.10). A phage-encoded molecular chaperone, gp57A, is required for the correct trimerisation of short tail fibre protein gp12 and the long tail fibre proteins gp34 and gp37. A specific chaperone, gp38, is required for the correct trimeric assembly of gp37. In the extended conformation, the tips of the long tail fibres can interact reversibly with receptor molecules. The glucosyl-α-1,3-glucose terminus of LPS or outer membrane porin C (OmpC) can function as alternative receptors and both outer membrane protein C and the interaction region of LPS have to be absent or inaccessible for *Escherichia coli* to be resistant to infection by T4.

When at least three long tail fibres have bound a receptor molecule, a signal is transferred to the baseplate of the phage, which then changes conformation. The binding information transfer is likely related to the angle of attachment of the long tail fibre to the baseplate. In the free phage, this angle is variable and the fibres are flexible up to certain limits. When several fibres are attached to their receptor, external forces on the phage may force these angles to less or more than these

allowed angles, pushing proteins near the long fibre attachment site in different conformations and triggering the conformational change of base plate, from the hexagon conformation to the star conformation [47]. This conformational change is very extensive, involving changes of location and interaction partners for several baseplate proteins. Among others, it leads to extension of the short tail fibres, which are normally folded away in the baseplate (Fig. 17.10b). The short tail fibres are trimers of the protein gp12. When the baseplate is in its hexagon conformation, the short tail fibres are incorporated into the baseplate by interactions along their lengths. However, when the baseplate switches to its star conformation, only the proximal (amino-terminal) end of gp12 remains bound to the baseplate from which the rest of the molecule extends. The distal carboxy-terminal part of the short tail fibres comprising residues 317–517 is responsible for a tight, irreversible interaction with the core region of the LPS (Fig. 17.10). The short tail fibres (gp12) have been seen in thin sections of phage-infected bacteria, forming rigid connections of the baseplate to the host cell surface.

In a subsequent step, contraction of the tail sheath (a helical polymer of protein gp18) is initiated [50]. The sheath of T4 is 24 nm wide and 93 nm long and is composed of 138 gp18 subunits, which are arranged in a six-start helix. The pitch and twist of the helix are 4 nm and 17°, respectively. The contracted T4 sheath is 33 nm wide and 42 nm long. The six-start helix changes its pitch and twist to 1.6 nm and 33°, respectively. In the process of sheath contraction, gp18 subunits move as rigid bodies without refolding or significant domain shifts. Upon contraction, the contact area between gp18 molecules increases by about four times.

As a result of tail sheath contraction, the rigid inner tail tube (a helical polymer of gp19) is driven through the outer cell membrane. The trimeric protein gp5 at the tip of the tail tube contains three lysozyme domains that are activated when they enter in contact with the periplasmic peptidoglycan layer. They create a hole, which allows reaching the cytoplasmic membrane. When the tail tube interacts with the cytoplasmic membrane, the phage DNA is released through the tail tube. As for other phages, the pressure of the tightly packed DNA in the head domain probably accounts for the force that initiates DNA transfer, but its completion may need further mechanisms. It appears a proton-motive force over the inner membrane is necessary for successful DNA transfer [52]. Under favourable conditions, the 172 kilobase genome of phage T4 DNA enters the bacterium is less than one minute, making it a very efficient process.

Bacteriophage P2. P2 is a myovirus with a smaller genome than that of T4, namely 33 kilobases [53]. It also has a more simple tail, for which the synthesis of only eleven gene products is necessary. Phage P2 sports a baseplate with a single central spike, which is a trimer of gpV, and six tail fibres, which are formed by gpG. Little is known about the structure of the baseplate, but the structure of the central spike protein has been solved at high resolution [48]. Like the λ side tail fibre and bacteriophage T4 distal tail fibre protein gp37, the fibres have a specialised chaperone to aid them in their folding. In the first step of infection, the fibres attach to the core region of the LPS. Calcium ions have been shown to greatly improve

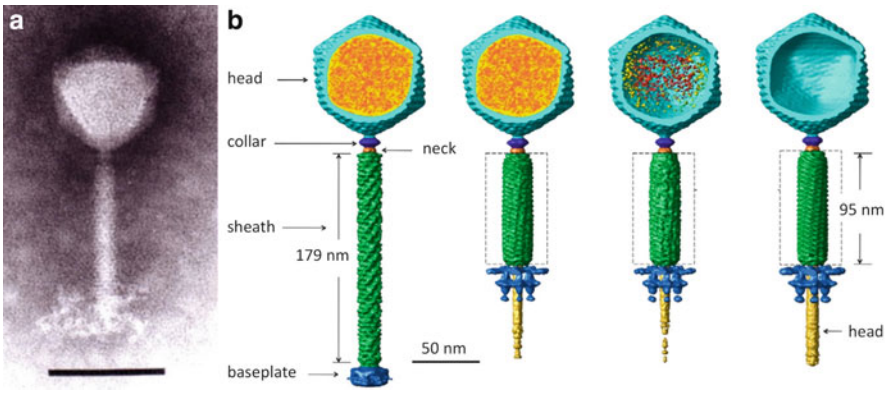


Fig. 17.11 Myoviruses of Gram-positive bacteria. (**a**) Electron micrograph of bacteriophage SPO1 (Reproduced from [54]. With permission). The complex baseplate is clearly visible, with multiple appendages that are presumably receptor-binding proteins. The bar indicates 100 nm. (**b**) *Bacillus anthracis* spore-binding phage 8a. Electron cryotomographic images of four stages on the pathway of DNA ejection are shown: the phage with extended sheath, the phage with contracted sheath but still containing DNA, the phage with contracted sheath in the process of ejecting its DNA and the phage with contracted sheath that has fully ejected all DNA (Reproduced from 55]. With permission). The conformational change of the baseplate from a disk-shaped structure to a structure with sideways- and downward-pointing extensions is very clear

adsorption. The baseplate then presumably changes conformation to allow the central tail tube to pass through it. Contraction of the tail sheath drives the tail tube through the baseplate and the bacterial outer membrane and periplasm, delivering the DNA exit site to the inner membrane. The tip of the tail tube consists of a triple-helical conical spike, of which the very tip is stabilised by three ions: iron, chloride and calcium (Fig. 17.10d).

Bacteriophage SPO1. An example of a myovirus that infects a Gram-positive bacterium is bacteriophage SPO1 [54] (Fig. 17.11a). SPO1 is a large, strictly virulent phage infecting *Bacillus subtilis*. It features a non-permuted double-stranded DNA genome of 146 kb, with redundant, invariable repeats of 13 kb at both ends. The long contractile tail consists of a complex baseplate and a sheath that is 140 nm long and 19 nm wide; the contracted sheath is 64 nm long and 27 nm wide. The baseplate, which functions as a receptor-recognition device and as a trigger for tail sheath contraction, undergoes structural rearrangement upon contraction, rearranging itself into a double-ringed structure. However, high-resolution studies on bacteriophage SPO1, its subcomplexes or proteins by cryo-electron microscopy or crystallography have not been performed.

Bacteriophage 8a. The structure of the *Bacillus anthracis* spore-binding phage 8a was determined by cryo-electron tomography (see Chap. 3) [55] (Fig.17.11b). This phage infects the Gram-positive *Bacillus anthracis* and, like the name suggests, it also binds *Bacillus anthracis* spores. Like other myoviruses, the tail consists of a six-

start helical sheath surrounding a central tail tube, but it has a baseplate different from other phages. The baseplate recognises and attaches to host cells. Cryo-electron tomographic analysis was performed on four different conformations that likely correspond to four states of the infection process: DNA-filled particles with extended tails, DNA-filled particles with contracted tails, partially DNA-filled particles with contracted tails and empty particles with contracted tails.

The baseplate has a claw-like structure, with six extensions in a closed conformation when the tail is extended, and in an open conformation when the tail is contracted. Primary receptor interaction probably takes place by the closed baseplate, although it is unknown which part of the baseplate is responsible, as phage 8a lacks long tail fibres or other obvious primary receptor binding modules. However, densities reminiscent of the bacteriophage T4 short tail fibres [47] are present at the periphery of the baseplate, in a folded-away state when the tail sheath is extended. Perhaps the folded-away short tail fibre-like densities present a sufficient receptor-binding site for interaction. In any case, the baseplate undergoes a conformational change from a closed to an open conformation upon binding to the host cell surface, because in the contracted state, these densities point downwards. The downward extension of the densities may bind host cell receptors more efficiently to irreversibly immobilise the baseplate on the host cell surface and cause its structural rearrangement. As the baseplate and tail sheath are physically connected, the structural rearrangement of the baseplate likely initiates contraction of the tail sheath.

The extended tail sheath is assembled around the 8 nm-diameter tail tube as a right-handed, six-stranded helical structure measuring 20 nm in diameter and 180 nm in length. Each helical strand consists of 45 subunits related to each other by a 4 nm helical pitch and a 22° twist. The 45 subunits in each helical strand of the contracted tail are related by a 2 nm pitch and a 31° twist; the contracted tail sheath is thus about half the length of the extended one. Like for bacteriophage T4, the structure of individual tail sheath subunits is similar in both conformations, indicating that tail contraction results from rigid body rotations and translations of entire subunits rather than by conformational changes within subunits. Contraction of the tail sheath forces the central tail tube downwards into the cell surface. A central spike structure is present at the distal end of the tail tube in extended tails, which may function as a cell-puncturing device as proposed for T4. No equivalent structure is seen in contracted tails, so it is likely the tail spike falls off during or after contraction to allow passage of the DNA.

Comparison of the neck regions in extended and contracted tails indicates that there are two conformations. In the extended state, the neck is assembled into a compact shape plugging the connection between the head and the tail tube. In the contracted state, the neck forms a central pore that is large enough to allow the passage of double stranded DNA. Thus, the neck protein may function as a gate that regulates the passage of DNA from the head into the tail, by undergoing a substantial conformational change upon tail contraction.

17.3 Perspectives and Conclusions

A bacteriophage that ejects its DNA spuriously or into an unsuitable host cell ceases to exist [31]. Evolution has therefore endowed bacteriophages with efficient host cell recognition and DNA transfer mechanisms. These mechanisms differ widely between phages, and it is therefore difficult to propose a general mechanism. Also, due to the research interests of the scientists studying particular phages, they have often concentrated on certain aspects of phage biology. Thus, much is known about the genetic regulation of phage λ, but the adsorption and DNA transfer processes have not been studied profoundly. On the other hand, the tail of bacteriophage T4 for example, has been analysed in great structural detail. However, given these restrictions, here we give a brief overview, attempt to draw some general conclusions and draw attention to some interesting phage facts and potential applications.

In general, bacteriophages attach to host cells using a two-step mechanism. In the first step, attachment is specific but reversible. This allows the phage to diffuse two-dimensionally over the bacterial surface until the second receptor is encountered. This second receptor is then bound irreversibly and at this point, the phage is immobilised and committed to DNA transfer. The sequences of receptor-binding proteins of phages are highly variable; and varying the receptor-binding properties is probably the main mechanism a phage has to adapt to hosts that have varied their cell wall structure or to adapt to a new bacterial host strain or species. Horizontal gene transfer between phages plays an important role in this variation, even between different phage types (horizontal gene transfer is probably most likely when a phage infects a bacterial cell containing a prophage). An example of relatively recent horizontal gene transfer is between the tip of the temperate siphovirus λ side tail fibre and the virulent myovirus T4 long tail fibre, in which residues that are structurally important are conserved, but the distal domain has evolved to bind a different receptor [46]. Another example of evolutionary recent horizontal gene transfer is the sequence and structural homology between the receptor-binding domains of the tailspikes of the podovirus P22 and the myovirus Det7 [56]. There are of course many phages still to be discovered, and their biological characterisation, the sequencing of their genomes and structure determination by electron microscopic (see Chap. 3) and crystallographic techniques (see Chap. 4) will likely reveal new receptor-recognition mechanisms.

Some phages encode several alternate tail fibre genes and may express them either alternately or simultaneously in the same particle. An example is the myovirus ϕ92, which encodes a Swiss army knife-like set of fibres to infect most laboratory strains of *Escherichia coli* (including those with a sialic acid capsule) and many *Salmonella* strains [57]. In the temperate myoviruses μ and P1 the specificity for the host receptor is altered upon inversion of a genome segment. The inversion results in a tail fibre with an alternative carboxy-terminal receptor-binding part. The amino-terminal part, which binds to the phage, is coded outside the inversion region and thus stays constant. In this way, in one form bacteriophage μ carries tail fibres that allow it to infect *Escherichia* and *Salmonella* bacteria and in a second form tail fibres that allow it to infect *Shigella*, *Enterobacter*, *Erwinia*, and *Citrobacter* [58]. *Bordetella* phage

BPP-1 has been shown to use an error-prone reverse transcriptase to generate variation in the distal knob of its tail fibre proteins [59].

As Gram-positive bacteria lack an outer membrane and instead have a thicker layer of peptidoglycan, one would expect their receptor-binding mechanisms to be different. Phages infecting Gram-negative hosts tend to use porins, transport proteins, enzymes or LPS available on the outer membrane as primary or secondary receptors. Phages infecting Gram-negative hosts tend to use peptidoglycan, teichoic acids or proteins that span the cell wall as receptors. One characteristic of tailed phages infecting Gram-positive hosts is that they appear to lack the side tail fibres many tailed phages infecting Gram-negative hosts possess. Phages may also bind to pili, flagella and capsules of both kinds of bacteria.

In most bacteriophages, much less is known about the mechanisms of nucleic acid transfer. Because the nucleic acid in phages is packaged under considerable pressure (see Chaps. 12, 18 and 19), it is likely initial transfer is at least helped by the partial release of this pressure when an appropriate transfer channel has been opened. Transfer of the remaining nucleic acid may then be promoted by phage proteins co-transferred into the cell, produced by transcription from the first region of nucleic acid transferred or by host factors. For a phage that always starts transfer with a specific end of the genome, like T7, proteins coded for at this end may be involved in successful genome transfer. For circularly permuted phages like T4, the entire nucleic acid is likely transferred by pressure and host factors, as the phage can not count on a specific protein being expressed from the part of the genome first transferred. Phages have also developed different mechanisms to protect the nucleic acid that is transferred into the cells. In general, nucleic acids are protected from nucleases in the periplasm by a physical barrier: the extensible tail tube of podoviruses and probably also siphoviruses, or the inner tail tube in myoviruses. Once in the cytoplasm, restriction enzymes are a major danger. Some phages modify the nucleotides of their DNA so they cannot be recognised by restriction enzymes. Other phages (T7, T5) only transfer a small part of their genome in a first step, and this part encodes proteins that inactivate restriction enzymes or otherwise protect DNA against restriction.

For nucleic acid transfer, phages infecting Gram-positive bacteria only have to cross the inner membrane. Therefore, if they can somehow get close to it by digesting the peptidoglycan cell wall, they may not need a to extend a tube like tailed phages infecting Gram-negative bacteria need to do to allow their genomes to safely traverse the nuclease-ridden periplasmic space. This may be an explanation for the fact that there appear to be relatively less myoviruses infecting Gram-positive bacteria [31]. Instead, podoviruses and siphoviruses can, at least in theory, digest their way straight to the cytoplasmic membrane of Gram-positive bacteria if their baseplates contain the appropriate enzymes. Alternatively, they can attach to the outer layer of the cell wall and extend a tube like phages that infect Gram-negative bacteria do.

Future structural studies of bacteriophage proteins, phage protein assemblies and whole bacteriophages by crystallography (Chap. 4) and electron microscopy (Chap. 3) will certainly resolve more details of phage receptor-binding and nucleic

acid transfer. Cryo-tomography (Chap. 3) using mini-cells of bacteria will play an important role in this, like performed for the *Escherichia coli* myovirus P1 [60]. Apart from structural studies, future studies will aim to measure individual interactions of phages and their proteins with their host cells. Single molecule techniques may be used to measure the force of binding of bacteriophage or their receptor-binding proteins to the receptor. In atomic force microscopy (see Chap. 8) a phage fibre may be bound to the tip and used to probe a surface containing its receptor, while optical tweezers (see Chap. 9) could also be used to measure these interactions.

The remarkable properties of bacteriophages have led to many ideas for applications (see Chap. 22). Now that information in atomic detail is becoming available on receptor-binding proteins of different bacteriophages, this information can be exploited to perform directed mutation and modulate their receptor-binding properties. Future knowledge of the nucleic transfer mechanisms may also lead to applications in designing nucleic acid delivery vehicles.

Acknowledgements Work in the authors' laboratory is supported by grant BFU2011-24843 from the Spanish Ministry of Economy and Competitiveness. Structure figures were made with PyMol (The PyMOL Molecular Graphics System, Schrödinger, LLC) or the UCSF Chimera package (developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco and supported by NIGMS 9P41GM103311).

References and Further Reading

1. Wommack KE, Colwell RR (2000) Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114
2. Marks R, Sharp R (2000) Bacteriophages and biotechnology: a review. J Chem Technol Biotechnol 75:6–17
3. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 185:60–89
4. Mesyanzhinov VV (2004) Bacteriophage T4: structure, assembly, and initiation infection studied in three dimensions. Adv Virus Res 63:287–352
5. Nakanishi M, Akuta T, Nagoshi E, Eguchi A, Mizuguchi H, Senda T (2001) Nuclear targeting of DNA. Eur J Pharm Sci 13:17–24
6. Tang J, Lander GC, Olia AS, Li R, Casjens S, Prevelige P Jr, Cingolani G, Baker TS, Johnson JE (2011) Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in p22. Structure 19:496–502
7. Specthrie L, Bullitt E, Horiuchi K, Model P, Russel M, Makowski L (1992) Construction of a microphage variant of filamentous bacteriophage. J Mol Biol 228:720–724
8. Toropova K, Basnak G, Twarock R, Stockley PG, Ranson NA (2008) The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. J Mol Biol 375:824–836
9. Bamford DH, Caldentey J, Bamford JK (1995) Bacteriophage PRD1: a broad host range dsDNA tectivirus with an internal membrane. Adv Virus Res 45:281–319
10. Butcher SJ, Manole V, Karhu NJ (2012) Lipid-containing viruses: bacteriophage PRD1 assembly. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York

11. Hayashi M, Aoyama A, Richardson DL, Hayashi MN (1988) Biology of the bacteriophage PhiX174. In: Calendar R (ed) The bacteriophages. Plenum Publishing Corporation, New York
12. Fane BA, Brentlinger KL, Burch AD, Hafenstein SL, Moore E, Novak CR, Uchiyama A (2005) PhiX174 et al., the microviridae. In: Calendar RL (ed) The bacteriophages, 2nd edn. Oxford University Press, Oxford
13. Golmohammadi R, Valegard K, Fridborg K, Liljas L (1993) The refined structure of bacteriophage MS2 at 2.8Å resolution. J Mol Biol 234:620–639
14. Toropova K, Stockley PG, Ranson NA (2011) Visualising a viral RNA genome poised for release from Its receptor complex. J Mol Biol 408:408–416
15. Riechmann L, Holliger P (1997) The C-terminal domain of TolA is the co-receptor for filamentous phage infection of *E. coli*. Cell 90:351–360
16. McKenna R, Ilag LL, Rossmann MG (1994) Analysis of the single-stranded DNA bacteriophage PhiX174, refined at a resolution of 3.0Å. J Mol Biol 237:517–543
17. Betts S, King J (1999) There's a right way and a wrong way: *in vivo* and *in vitro* folding, misfolding and subunit assembly of the P22 tailspike. Structure 7:R131–R139
18. Olia AS, Casjens S, Cingolani G (2007) Structure of phage P22 cell envelope–penetrating needle. Nat Struct Mol Biol 14:1221–1226
19. Casjens SR, Molineux IJ (2012) Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York
20. Chang JT, Schmid MF, Haase-Pettingell C, Weigele PR, King JA, Chiu W (2010) Visualizing the structural changes of bacteriophage Epsilon15 and its *Salmonella* host during infection. J Mol Biol 402:731–740
21. Garcia-Doval C, van Raaij MJ (2012) Crystallization of the C-terminal domain of the bacteriophage T7 fibre protein gp17. Acta Crystallogr Sect F Struct Biol Cryst Commun 68:166–171
22. Steven AC, Trus BL, Maizel JV, Unser M, Parry DA, Wall JS, Hainfield JF, Studier FW (1988) Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J Mol Biol 200:351–365
23. Garcia-Doval C, van Raaij (2012) Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. Proc Natl Acad Sci USA 109:9390–9395
24. Chang CY, Kemp P, Molineux IJ (2010) Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. Virology 398:176–186
25. Xiang Y, Rossmann MG (2011) Structure of bacteriophage Phi29 head fibers has a supercoiled triple repeating helix-turn-helix motif. Proc Natl Acad Sci USA 108:4806–4810
26. Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, Rossmann MG (2009) Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Mol Cell 34:375–386
27. Xiang Y, Morais MC, Battisti AJ, Grimes S, Jardine PJ, Anderson DL, Rossmann MG (2006) Structural changes of bacteriophage Phi29 upon DNA packaging and release. EMBO J 25:5229–5239
28. Xiang Y, Morais MC, Cohen DN, Bowman VD, Anderson DL, Rossmann MG (2008) Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage Phi29 tail. Proc Natl Acad Sci USA 105:9552–9557
29. Alcorlo M, Gonzalez-Huici V, Hermoso JM, Meijer WJ, Salas M (2007) The phage Phi29 membrane protein p16.7, involved in DNA replication, is required for efficient ejection of the viral genome. J Bacteriol 189:5542–5549
30. Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV (2007) Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J 26:3720–3728
31. Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL (2012) Long noncontractile tail machines of bacteriophages. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York

32. Letellier L, Boulanger P, Plancon L, Jacquot P, Santamaria M (2004) Main features on tailed phage, host recognition and DNA uptake. Front Biosci 9:1228–1339
33. Heller K, Braun V (1979) Accelerated adsorption of bacteriophage T5 to *Escherichia coli* F, resulting from reversible tail fiber-lipopolysaccharide binding. J Bacteriol 139:32–38
34. Heller K, Braun V (1982) Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5 *via* the L-shaped tail fibers. J Virol 41:222–227
35. Boehm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL (2001) FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol 11:1168–1175
36. Guihard G, Boulanger P, Letellier L (1992) Involvement of phage T5 tail proteins and contact sites between the outer and inner membrane of *Escherichia coli* in phage T5 DNA injection. J Biol Chem 267:3173–3178
37. Hendrix RW, Duda RL (1992) Bacteriophage Lambda PaPa: not the mother of all Lambda phages. Science 258:1145–1148
38. Montag D, Schwarz H, Henning U (1989) A component of the side tail fiber of Escherichia coli bacteriophage Lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. J Bacteriol 171:4378–4384
39. Berkane E, Orlik F, Stegmeier JF, Charbit A, Winterhalter M, Benz R (2006) Interaction of bacteriophage Lambda with its cell surface receptor: an *in vitro* study of binding of the viral tail protein gpJ to LamB (maltoporin). Biochemistry 45:2708–2720
40. Roessner CA, Ihler GM (1986) Formation of transmembrane channels in liposomes during injection of Lambda DNA. J Biol Chem 261:386–390
41. Erni B, Zanolari B, Kocher HP (1987) The mannose permease of *Escherichia coli* consists of three different proteins. Amino acid sequence and function in sugar transport, sugar phosphorylation, and penetration of phage Lambda DNA. J Biol Chem 262:5238–5247
42. Veesler D, Robin G, Lichiere J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C (2010) Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in Gram-positive infecting phages. J Biol Chem 285:36666–36673
43. Lhuillier S, Gallopin M, Gilquin B, Brasiles S, Lancelot N, Letellier G, Gilles M, Dethan G, Orlova EV, Couprie J, Tavares P, Zinn-Justin S (2009) Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc Natl Acad Sci USA 106:8507–8512
44. Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichière J, van Heel M, Campanacci V, Moineau S, Cambillau C (2010) Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci U S A 107:6852–6857
45. Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C (2012) Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A 109:8954–8958
46. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, van Raaij MJ (2010) Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proc Natl Acad Sci USA 107:20287–20292
47. Leiman PG, Arisaka F, van Raaij MJ, Kostyuchenko VA, Aksyuk AA, Kanamaru S, Rossmann MG (2010) Morphogenesis of the T4 tail and tail fibers. Virol J 7:355
48. Browning C, Shneider MM, Bowman VD, Schwarzer D, Leiman PG (2012) Phage pierces the host cell membrane with the iron-loaded spike. Structure 20:326–339
49. Rao VB, Black LW (2010) Structure and assembly of bacteriophage T4 head. Virol J 7:356
50. Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. In: Rossmann MG, Rao V (eds) Viral molecular machines. Springer, New York
51. Wood WB, Eiserling FA, Crowther RA (1994) Long tail fibers: genes, proteins, structure, and assembly. In: Karam JD (ed) Molecular biology of bacteriophage T4. ASM Press, Washington, DC

52. Goldberg EB, Grinius L, Letellier L (1994) Recognition, attachment and injection. In: Karam JD (ed) Molecular biology of bacteriophage T4. ASM Press, Washington, DC
53. Nilsson AS, Haggard-Ljungquist E (2005) The P2-like bacteriophage. In: Calendar RL (ed) The bacteriophages, 2nd edn. Oxford University Press, Oxford
54. Klumpp J, Lavigne R, Loesnner MJ, Ackermann HW (2010) The SPO1-related bacteriophages. Arch Virol 155:1547–1561
55. Fu X, Walter MH, Paredes A, Morais MC, Liu J (2011) The mechanism of DNA ejection in the *Bacillus anthracis* spore-binding phage 8a revealed by cryo-electron tomography. Virology 421:141–148
56. Walter M, Fiedler C, Grassl R, Biebl M, Rachel R, Hermo-Parrado XL, Llamas-Saiz AL, Seckler R, Miller S, van Raaij MJ (2008) Structure of the receptor-binding protein of bacteriophage Det7: a podoviral tail spike in a myovirus. J Virol 82:2265–2273
57. Schwarzer D, Buettner FF, Browning C, Nazarov S, Rabsch W, Bethe A, Oberbeck A, Bowman VD, Stummeyer K, Muhlenhoff M, Leiman PG, Gerardy-Schahn R (2012) A multivalent adsorption apparatus explains the broad host range of phage Phi92: a comprehensive genomic and structural analysis. J Virol 86:10384–10398
58. Paolozzi L, Ghelardini P (2006) The bacteriophage Mu. In: Calendar RL (ed) The bacteriophages, 2nd edn. Oxford University Press, Oxford
59. Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, Miller JF (2002) Reverse-transcriptase mediated tropism switching in *Bordetella* bacteriophage. Science 295:2091–2094
60. Liu J, Chen CY, Siomi D, Niki H, Margolin W (2011) Visualization of bacteriophage P1 infection by cryo-electron tomography of tiny *Escherichia coli*. Virology 417:304–311

Further Reading

Calendar RL (2005) The bacteriophages, 2nd edn. Oxford University Press, Oxford
Campbell AM (1996) Bacteriophages. In: Fields BN, Knipe DM, Howley PM (eds) Fields virology, 3rd edn. Lippincott Williams & Wilkins, Philadelphia
Karam JD (1994) Molecular biology of bacteriophage T4. ASM Press, Washington, DC
Karam JD, Miller ES (2010) Bacteriophage T4 and its relatives: a series of critical reviews. Biomed Central, London
Rakhuba DV, Kolomiets EI, Dey ES, Novik GI (2010) Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. Pol J Microbiol 59:145–155
Rossmann MG, Rao VB (2012) Viral molecular machines. Adv Exp Med Biol, vol 726, Springer, New York

Also especially recommended for further reading are references [2, 3, 4, 9, 11, 17, 32, 54, 58] *listed above.*

Chapter 18
Mechanical Properties of Viruses

Pedro J. de Pablo and Mauricio G. Mateu

Abstract Structural biology techniques have greatly contributed to unveil the relationships between structure, properties and functions of viruses. In recent years, classic structural approaches are being complemented by single-molecule techniques such as atomic force microscopy and optical tweezers to study physical properties and functions of viral particles that are not accessible to classic structural techniques. Among these features are mechanical properties such as stiffness, intrinsic elasticity, tensile strength and material fatigue. The field of virus mechanics is contributing to materials science by investigating some physical parameters of "soft" biological matter and biological nano-objects. Virus mechanics studies are also starting to unveil the biological implications of physical properties of viruses. Growing evidence indicate that viruses are subjected to internal and external forces, and that they may have adapted to withstand and even use those forces. This chapter describes what is known on the mechanical properties of virus particles, their structural determinants, and possible biological implications, of which several examples are provided.

Keywords Virus • Capsid • Atomic Force Microscopy • Mechanical Properties • Elasticity • Stiffness • Spring Constant • Tensile Strength • Material Fatigue • Molecular Structure • Stability • Dynamics • Biological Materials • Virus Engineering • Biotechnology • Nanotechnology

P.J. de Pablo (✉)
Department of Physics of the Condensed Matter,
C03, Facultad de Ciencias, Universidad Autónoma de Madrid,
Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: p.j.depablo@uam.es

M.G. Mateu (✉)
Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM) and Department of Molecular Biology of the Universidad Autónoma de Madrid,
c/Nicolás Cabrera 1, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: mgarcia@cbm.uam.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_18,

Abbreviations

AFM	Atomic force microscopy
CCMV	Cowpea chlorotic mottle virus
ds	Double-stranded
EM	Electron microscopy
FEA	Finite-element analysis
FZ	Force *vs*. z-piezo displacement
HBV	Hepatitis B virus
HIV-1	Human immunodeficiency virus type 1
HSV-1	Herpes simplex virus type 1
MLV	Murine leukemia virus
MVM	Minute virus of mice
NV	Norovirus (Norwalk virus)
ss	Single-stranded
vdW	Van der Waals

18.1 Introduction

The mechanical properties of some biomolecular complexes are essential for their function. Indeed, forces at the nanoscale play a central role in biochemistry, from the myosin-actin system [1], which is the ultimate responsible of muscular action, to cellular or viral motor protein assemblies ([2], see Chaps. 9 and 12).

Viruses are striking examples of biomolecular complexes endorsed with specific material properties that may provide a basis to understand some aspects of their biophysical function, and their ability to endure a variety of environmental aggressions. Indeed, during the passive extracellular stage of the infection cycle the viral capsid can be considered as a container [3] (see Chap. 2) that protects the viral genome against physicochemical assaults [4]. This critical function of many viral capsids have imposed strong selection pressures on them, leading to structural stabilization. On the other hand, the need to uncoat the viral genome during infection of host cells has favored the evolution of many capsids as metastable, conformationally dynamic biological complexes. Thus, nature may have modulated some of the capsid material characteristics for both structural stability and dynamics (see Chap. 1). Different examples of the delicate balance between viral capsid structural stability and dynamics and its control during the viral cycle are provided in Chaps. 1, 10, 11, 13, 15 and 19.

Mechanical properties such as elasticity or tensile strength are among the material properties of any solid object including virus particles. The development of atomic force microscopy (AFM) (see [5] and Chap. 8) has enabled the study of the

mechanical properties of virions and virus capsids at the single-particle level [6]. In this chapter we describe current approaches to determine mechanical properties of virus particles that are inherent to materials science analyses. These properties include particle stiffness (by measuring spring constant values), intrinsic elasticity of the capsid material (Young's Modulus), brittleness and material fatigue. We then discuss the relationships between these mechanical properties and the molecular structure of viruses, possible biological implications and bio/nanotechnological applications of virus mechanics.

18.2 Mechanical Stiffness of Virus Particles Determined by AFM in Indentation Assays

The advent of single-molecule techniques, such as AFM (Chap. 8), optical tweezers (Chap. 9) and similar methods [7] opened the possibility of measuring the tiny forces (nN or pN) that take place in biological molecules and processes at the nanometer scale, such as conformational changes [8], protein folding [9] or the mechanochemical action of motor proteins [10] (see Chaps. 9 and 12). In particular, AFM is ideally suited to probe the mechanical elasticity of biological specimens in liquid, in close to physiological conditions (see also Chap. 8). AFM was applied early to cells [11] and microtubules [12], and it was just a matter of (a short) time that it was also applied to study virus elasticity. It may be illustrative to briefly refer here the origin of the first experiments on the mechanical properties of small, spherical virus particles in liquid using phage ϕ29 as a model [6]. Optical tweezers had been used to study the mechanochemical action of the ϕ29 DNA packaging motor [13] (see Chap. 9), and other experiments using optical tweezers on this same phage were later started by colleagues of one of the authors of this chapter (P.J.P.). Since these experiments were not working properly, we decided to check the virus integrity by AFM imaging. From that moment on, it was obvious that the elasticity experiments we and others had carried out by indenting (much larger) cells or (much longer) rod-like microtubules with the AFM tip could be readily adapted to quantitatively probe the mechanical features of even the smallest spherical viruses.

Chapter 8 describes the technique of AFM and some experimental setups regarding its use for virus imaging. In this section we will briefly describe some particular conditions required for measuring the mechanical stiffness of viral particles using AFM. The procedure requires first the strong adsorption of the specimens on a suitable surface, in order to minimize unwanted displacements during relatively deep indentations. In the case of viruses, a typical recipe is to silanize glass surfaces to enhance the hydrophobic interactions with the viral particles [6]. The virus suspension is diluted to get a few particles (3–5) per square micron on the surface and guarantee their convenient localization, while avoiding crowding (Fig. 18.1a). Jumping mode [14] has shown to be extremely successful for imaging individual viruses on surfaces (see Chap. 8). It is convenient to prewet the

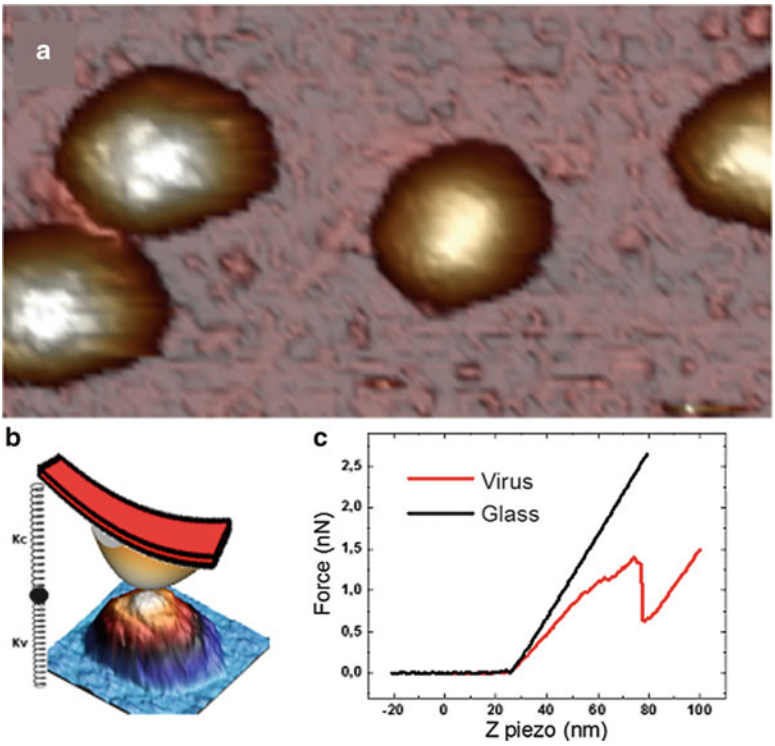


Fig. 18.1 Single indentation assay. (**a**) AFM topography of ϕ29 phage particles adsorbed on HOPG. (**b**) a sketch of an ideal nanoindentation of a viral particle and the interpretation of the mechanical result by considering the AFM cantilever and the virus particle as a system of two springs in series. **c** FZ (force *vs*. z-piezo displacement) curves on a hard substrate (*dark*) and a virus particle (*red*)

tip with ~20 μl of buffer to avoid undesired cantilever bending. A great amount of the work published on virus mechanics has been performed with AFM systems provided by *Nanotec Electrónica* [15]. Rectangular cantilevers RC800PSA, and Biolevers (BL-AC40TS) (*Olympus*) with nominal spring constants of 0.05 N/m and 0.03 N/m, respectively, are typically used. Cantilevers spring constants are typically calibrated using Sader's method [16]. Siffness determination is based in the monitored indentation of individual viral particles with the AFM tip. Therefore, in order to perform nano-indentations, single FZ (force *vs*. z-piezo displacement) experiments are performed by *pushing* on the top of a selected single virus particle. The particle is continuously zoomed in by reducing the x-y scanning size until the bump of the very top is under the whole piezo scan (about ~50 nm × 50 nm). Afterwards the FZ is executed by indenting the top of the particle, likely within a few nm of uncertainty (mainly provoked by thermal drift and the intrinsic non-linearity and creep of the piezo). This method has proven to be robust enough to

establish electrical contact with carbon nanotubes [17], which are much smaller in diameter than viral particles.

When the tip is in contact with the sample the tip-sample force F can be described by:

$$F = F_{contact} + F_{DLVO}$$

$F_{contact}$ is determined by the contact mechanics. In general, the expression of this force is a complicated solution of the elasticity equations, but there are two special and useful situations where it adopts a simple form:

i. *Hertzian contact* – for the gentle indentation of a spherical tip on an elastic solid half-space,

$$F_{contact} = 4/3E^* \sqrt{R} d^{3/2}$$

where R is the tip radius, d the indentation depth, and E^* is given by:

$$1/E^* = \left(1 - {\nu_t}^2\right)/E_t + \left(1 - {\nu_s}^2\right)/E_s$$

being ν_t, E_t, ν_s and E_s the Poisson's ratio and elastic modulus of the tip (t) and the sample (s) (defined below).

ii. *Linear contact* – when the effective response is like that of a spring with a constant k_{eff}. This situation occurs for small indentations of a thin shell-like structure [18], and

$$F_{contact} = -k_{eff} d$$

F_{DLVO} follows the Derjaguin-Landau-Verwey-Overbeek model (DLVO) [19] and accounts for the double Debye double-layer electrostatic force (F_{el}) and van der Waals (vdW) force (F_{vdW}). In the simplest case of two planes the forces are given by:

$$F_{DLVO} = F_{el} + F_{vdW} = \frac{2\sigma_s \sigma_p}{\varepsilon_0 \varepsilon} e^{\frac{-z}{\lambda_D}} - \frac{H_a}{6\pi z^3}$$

where σ_s and σ_t are the charge density of sample and tip, respectively; $\varepsilon\varepsilon_0$ is the dielectric constant times the permittivity of vacuum; z is the tip-sample distance, H_a is the Hammaker constant describing the van der Waals interactions between tip and sample, and λ_D is the Debye length that determines the range of the electrostatic forces.

We must consider two issues at this point. First, it is evident that to prime the measurement of sample elasticity it is necessary to decrease the Debye length. Since the Debye length decreases with the concentration c of salt in solution as $\lambda_D \sim 1/\sqrt{c}$ [20], when investigating viruses relatively high (but physiological) salt

concentrations are usually employed to screen as much as possible the electrostatic forces. In those circumstances, the tip-sample interactions are dominated by the contact forces and sample elasticity (Fig. 18.1b). Second, since in a first approximation most icosahedral virus capsids can be considered as hollow, relatively thin shells, it is tempting to use directly the linear approximation for modeling the tip-virus contact. Nevertheless, some caution is necessary because of two main reasons. On one hand, the linear model can only be applied if the total indentation of the virus capsid is at most the shell thickness (see [18] and Fig. 18.1c), which in most viruses is about a couple of nanometers. On the other hand, capsids or virions may be filled with either scaffolding proteins or nucleic acid, thus deviating from an ideal shell-like model. Even so, since these fillings cannot be considered as homogeneous solids in order to use the Hertzian contact approach, the linear model is still approximately valid and provides an easy and direct framework to interpret the experimental results.

In most AFM nanoindentation experiments on viruses it has been found that, for small deformations, the force is indeed linear with the indentation. In such cases, the spring constant of the virus k_v can be easily obtained by simply considering the virus like a spring in series with the cantilever (Fig. 18.1b):

$$k_v^{-1} = k_{eff}^{-1} - k_c^{-1}$$

where k_c is the spring constant of the cantilever (here the substrate is considered as non-deformable). For practical use, this equation is customarily rewritten in terms of the slope of the cantilever deflection on the glass, s_g (nm/V), and the slope of the cantilever deflection on the virus s_v (nm/V), as:

$$k_v = k_c \frac{s_g}{s_v - s_g}$$

A standard protocol of an AFM nanoindentation experiment is as follows. Each viral particle is indented with a sequence of about 5 FZ's. Following each FZ set, an image of the virus is taken to confirm its integrity, by checking that several effects, such as damage [21], collapse [12], or buckling [22] have not occurred; the image serves also to ascertain the particle position in order to correct for any drift, if needed, before performing the next FZ set. It is convenient to adjust the FZ speed to about ~60 nm/s in order to allow the water leave the virus when it is squeezed [23]. Even if the shell integrity is maintained, only those viral particles showing stable values for the spring constant along successive FZ measurements should be considered. Thus artifacts such as particle mobility effects that often occur when the particle is loosely bound to the surface can be mostly avoided. However, effects such as tip sliding or particle tilting are difficult to quantify, and techniques to rule them out remain to be developed. On the other hand, if breakage of the virus particle [24] is desired to determine tensile strength (rather than particle stiffness) the z-piezo elongation should be large enough to provoke a cantilever deflection similar

to the breaking force [6]. Several particles are normally used in the experiments, and the results are appropriately averaged.

It may be tempting to directly relate virus particle stiffness with structural stability against disruption into subunits. This interpretation may imply that "soft" virus particles are less resistant to dissociation than "hard" ones. However, very often elastic objects are more resilient than rigid ones. We can find examples in everyday life, *e.g.*, rubber *vs.* glass. Rubber is softer than glass, but glass do not support mechanical deformation without breaking as rubber does. Thus, other parameters such as tensile strength of virus particles need to be investigated.

18.3 Intrinsic Elasticity of Virus Capsids: Young's Modulus

In materials science, the Young's modulus of elasticity E provides a measure of the *intrinsic* stiffness of an elastic material. Let us first consider a solid piece of some material, say a cylinder of initial length L_0 and a cross-section with a surface A. If we apply a force F along the main axis, the cylinder elongates from its original length L_0 to $L_0+\Delta\ell$. For small deformations, the stress $\sigma = F/A$, *i.e.* the force per unit area A, is proportional to the relative deformation or strain ε, given by $\Delta\ell/L_0$ (Fig. 18.2a). Thus, $\sigma = E\varepsilon$, where the proportionality constant E is the Young's modulus. From the latter we can find an expression for the force [25]

$$F = \frac{AE}{L_0}\Delta\ell$$

The coefficient of $\Delta\ell$ has units of N/m, being the effective longitudinal spring constant k of the object (see Fig. 18.2a). Other important mechanical magnitude is the non-dimensional Poisson ratio parameter μ, which accounts for the ratio of transverse and longitudinal strains:

$$\mu = -\frac{\varepsilon_{trans}}{\varepsilon_{long}}$$

Transversal strain ε_{trans} is typically negative, since when a piece of material is pulled by the force F the transversal dimension decreases by Δw (Fig. 18.2b). There are exceptions like auxetic materials that, when stretched, become thicker perpendicularly to the applied force, and thus have negative values of the Poisson ratio. In general, the Poisson ratio μ value is constrained between -1 and 0.5, and steel and rigid polymers show values around 0.3 [26]. Actually, in studies of virus mechanics μ is assumed to be 0.3.

In the following we will go through the typical approaches followed in most published works to extract the Young's modulus of viral shells, *i.e.*, thin-shell theory and finite element analysis (see also Chap. 19). Certainly, from a physical

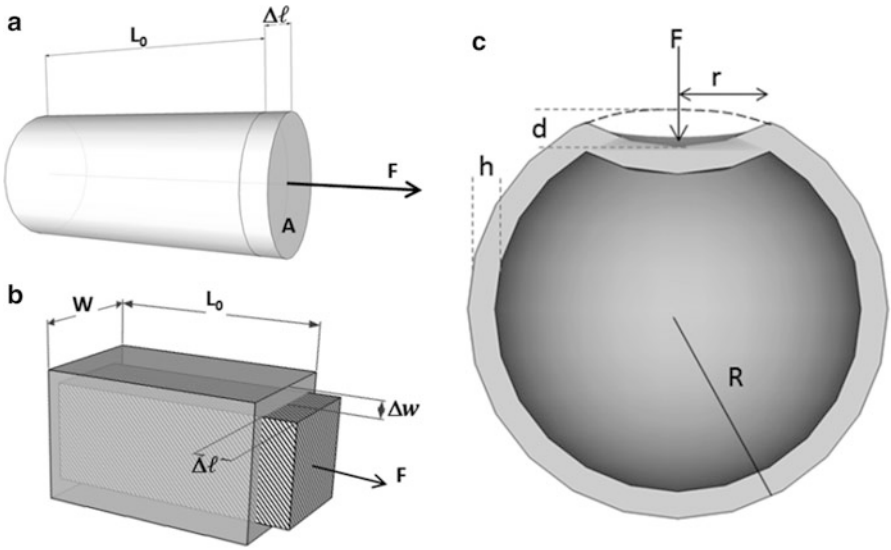


Fig. 18.2 Physical magnitudes of materials mechanics. (**a**, **b**) Geometrical parameters required for estimating the Young's Modulus (**a**) or the Poisson Ratio (**b**). (**c**) Main geometrical definitions for the application of thin shell theory (see text)

point of view, one outcome of most works published on virus mechanics so far is the estimation of the Young's modulus of a virus shell.

18.3.1 Shell Theory

It is well known that shells, *i.e.* thin plates which are curved in their undeformed state, exhibit special mechanical properties. For the sake of clarity, let us begin with a thin and straight rod. When supported at both ends, transverse forces cause bending and deflection. However, the same rod can withstand much greater axial load with unnoticeable deformation. Similar concepts apply to a thin flat shell. Perpendicular forces actuating on a flat plate would generate bending, resisted by bending moments in the cross-section. In-plane forces, in turn, would cause stretching, compression and/or shearing, resisted by in-plane forces on a cross-section. A closed curved shell would be subjected to both bending and stretching under the application of a perpendicular force, offering more mechanical strength than the flat plate that will essentially be bent without stretching. In the following we will analyze the details of these ideas in thin curved shells. When a spherical shell of radius R and thickness $h \ll R$ is deformed a distance d by an indenting point force F in the regime where $d \ll R$, the shell is locally both stretched and bent (Fig. 18.2c) [18]. As a consequence a bulge of size r develops right at the

application point of F. We can express the stretching and bending energies as [27]:

$$E_{str} \sim K\left(\frac{d}{R}\right)^2 r^2, \text{ and } E_b \sim \kappa\left(\frac{d}{r^2}\right)^2 r^2$$

where $K = \frac{Eh}{2(1-\mu)}$ and $\kappa = \frac{Eh^3}{12(1-\mu^2)}$ are the 2D stretching and bending modulus, respectively. Note that these refer to the three-dimensional properties of the material making up the shell. We can write the sum of bending and stretching energies as:

$$E_{total} \sim E\frac{h^2}{R}\left[\frac{hR}{r^2} + \frac{r^2}{hR}\right]d^2$$

By imposing the equilibrium condition $\frac{dE_{total}}{dr} = 0$ we can find that the size of the bent region scales as $r \sim \sqrt{hR}$. This means, for instance, that for two spherical shells with identical thickness h, doubling the radius R only increases the lateral deformation by a factor $\sqrt{2}$. By making the derivative of E_{total} with respect to the indentation d and using the obtained value of r, we find that the force is linear in the indentation with a spring constant of the shell that scales like $\sim E\frac{h^2}{R}$. To calculate the proportionality constant accurately requires to expand the shape of the deformed shell in spherical harmonics [12, 28], and for a thin shell it turns out to be $\frac{2}{\sqrt{3(1-\nu^2)}}$.

18.3.2 Finite Element Analysis

Another popular approximation for interpreting the results of experiments on virus mechanics is to use Finite Element Analysis (FEA). In simple terms, FEA analysis attempts to obtain a numerical solution of the complicated differential equations describing a physical problem concerning a physical object considered as a continuous medium [29]. The key-point is to divide the object in many small non-intersecting domains (finite elements) where the physical equations are discretized and solved. This numerical methodology can be applied for a variety of physical phenomena, including mechanical deformations, circulation of fluids or electrostatic problems.

In order to simulate the mechanical deformation of a virus capsid, an idealized, simplified geometrical model of the capsid is obtained from the three-dimensional structure provided by X-ray crystallography or cryo-electron microscopy (cryo-EM) (Fig. 18.3a, b) using commercially available software. Young's modulus and Poisson ratio values are established as material properties of the model. Afterwards, the proper boundary conditions are fixed and the FEA structure is deformed by punctual forces, Hertzian forces or spheres (Fig. 18.3c). The values of the forces used in FEA

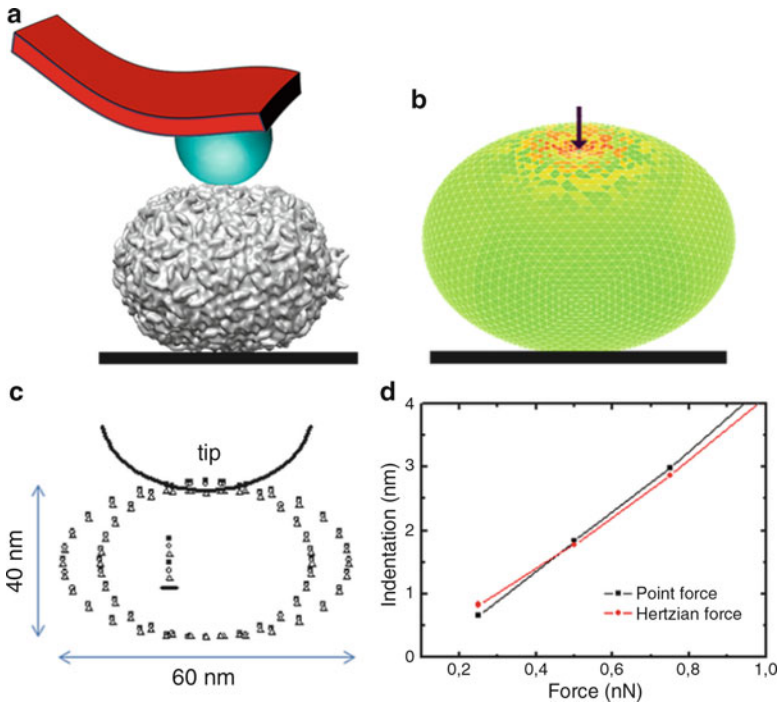


Fig. 18.3 Finite Elements Modeling. (**a**) Cartoon that depicts the actual geometry of a simple AFM tip and a morphologically complex virus (phage ϕ29). (**b**) Simplified FEA modeling of phage ϕ29 as used in ref. [6]. (**c**) Maximum and minimum cross-section variations of the FEA model as the tip deforms the virus model. (**d**) Typical FEA results of indentations by using point and Hertzian forces as a function of the indenting force

are the same than those actually used in AFM nanoindentation experiments. Therefore, deformation of the model is obtained as a function of the loading force and the indentation is depicted *vs.* force. Interestingly, for indentations of the order of the shell thickness, punctual and Hertzian forces result in similar deformations (Fig. 18.3d). Thus the typical approach is to vary the Young's modulus to match the experimental indentation in actual AFM experiments. Specifically, FEA models of viruses are designed in terms of the capsid general shape, diameter and thickness. However, it is difficult to capture the complexity of a virus capsid structure with FEA. The proteins forming the capsid and the bonds between them cannot be modeled just by using Young's modulus and Poisson ratio values. Nonetheless, FEA provides a coarse approach for understanding the mechanical properties of the capsid protein material [6, 30]. In order to refine the calculation, some studies have performed FEA using virus models that approach more closely cryo-EM structural models [21]. Other studies have contemplated coarse grained models that include as discrete pieces the capsid subunits or capsomers that form the capsid structures, and approximate the interactions between subunits using harmonic or Lennard Jones potentials [31].

18.4 Brittleness and Material Fatigue of Virus Particles

A material is brittle if it breaks without a significant deformation (strain) when subjected to mechanical deformation. This is the case of glass or an eggshell. Brittle materials usually do not present plastic deformation before breaking. On the other hand, material fatigue studies apply cycling loading forces below rupture strength of the object under examination [32], revealing not only the evolution with time and repeated application of force of its resistance to breakage, but also the location of the object's weakest parts. A fundamental requirement of these studies is monitoring the structural integrity of the sample along the experiment. The investigation of both mechanical characteristics of virus particles require a precise control of the force applied on the particle [33] to induce sequential disruption of the capsid subunits. These novel mechanical studies on viral particles may provide biologically relevant information about virus disassembly.

18.5 A Case Study: The Mechanics of Bacteriophage ϕ29

In this section we will exemplify the combination of AFM experiments and FEA modeling to extract some biophysical implications on the mechanical properties of virus particles, using bacteriophage ϕ29 as a model. First, we will discuss the values obtained for the spring constant of the ϕ29 prohead, or immature capsid (as a measure of capsid mechanical stiffness), and its relationship with the Young's modulus (as a measure of the intrinsic elasticity of the capsid material) by using Finite Element modeling. Second, we will explain the detection and evaluation of internal pressure in bacteriophage ϕ29. Finally, we will describe the identification of built-in mechanical stress in viral particles and its influence on capsid stiffness.

18.5.1 Mechanical Properties of the ϕ29 Capsid: Stiffness and Structural Stability Against Breakage

Indentation assays of individual ϕ29 proheads revealed the existence of a bimodal distribution of the spring constant [6], and yielded average values of 0.16 N/m and 0.31 N/m. However, further studies of the same proheads [31, 34] revealed a single distribution of the spring constant with an average value of about 0.3 N/m. It is possible to calculate analytically the deformation of a homogeneous, spherical, and elastic shell that is subjected to equal and opposite forces applied at the poles by expanding the shape of the deformed sphere in spherical harmonics [28]. The spring constant for the total indentation becomes $k = 2.25\frac{Eh^2}{R}$, and the Young's modulus of the shell can be estimated, by using the thickness h and the radius R of the phage particle (obtained from EM studies [35]), to be between 1.2 GPa and 1.6 GPa.

Further modeling with FEA (Sect. 18.3.2) results in a Young's modulus of 1.8 GPa. Phage ϕ29 proheads also showed a decrease of the spring constant after repeated indentations. Further experiments suggested that this capsid softening was probably due to capsid disruption [24, 36].

18.5.2 Internal Pressure in the ϕ29 Virion

It has been proposed that translocation of the double-stranded (ds) DNA genome of phage ϕ29 through the virus tail into the bacterium host is initiated by a *push* mechanism, followed by a *pull* mechanism mediated by bacterium proteins [37]. The push mechanism would be driven by the elastic energy provided by the internal pressure accumulated during the DNA packaging process (see also Chaps. 17 and 19). However, the very existence of pressure in ϕ29 has remained elusive, and single-molecule techniques have added valuable information. Specifically, AFM nanoindentation experiments on individual viral particles reveal that the ϕ29 virion is stiffer than the ϕ29 prohead (Fig. 18.4b). Since no major mechanically-relevant structural differences between the prohead and the empty mature head particles are apparent, it is reasonable to consider that the packaged dsDNA is the main responsible of the observed stiffening.

In order to distinguish between structural or pressure-induced reinforcement, further experiments were performed in the presence of spermidine^{3+}, a positively charged counter ion that can induce DNA condensation, even inside viruses [38], because it neutralizes the negative charges of the nucleic acid phosphates. Real-time indentation experiments probing the same individual particles before and during treatment with spermidine^{3+} showed that this compound induces a mechanical softening of the full virion that is made as elastic as the prohead, as determined by measuring spring constant values. Experiments carried out by adding and later removing spermidine^{3+} showed that this is a reversible process (Fig. 18.4c). Theoretical calculations using the model described in [39] yielded an internal pressure in the ϕ29 virion of about 30 ± 10 atm in those conditions, in the absence of spermidine^{3+}. The experimental evaluation of the internal pressure from the determination of spring constant values in the absence or presence of spermidine^{3+} and the use of FEA (see Sect. 18.3.2 and Chap. 19) yielded a pressure of 40 ± 10 atm, in good agreement with the theoretical predictions (Fig. 18.4d) [34].

18.5.3 Built-in Mechanical Stress in the ϕ29 Capsid

We have already commented on the mechanical advantage that curved shells present when subjected to deformation (*e.g.*, by indentation), as they undergo simultaneous stretching and bending. Nevertheless, there are other mechanisms which viral shells may take advantage of to improve their mechanical performance.

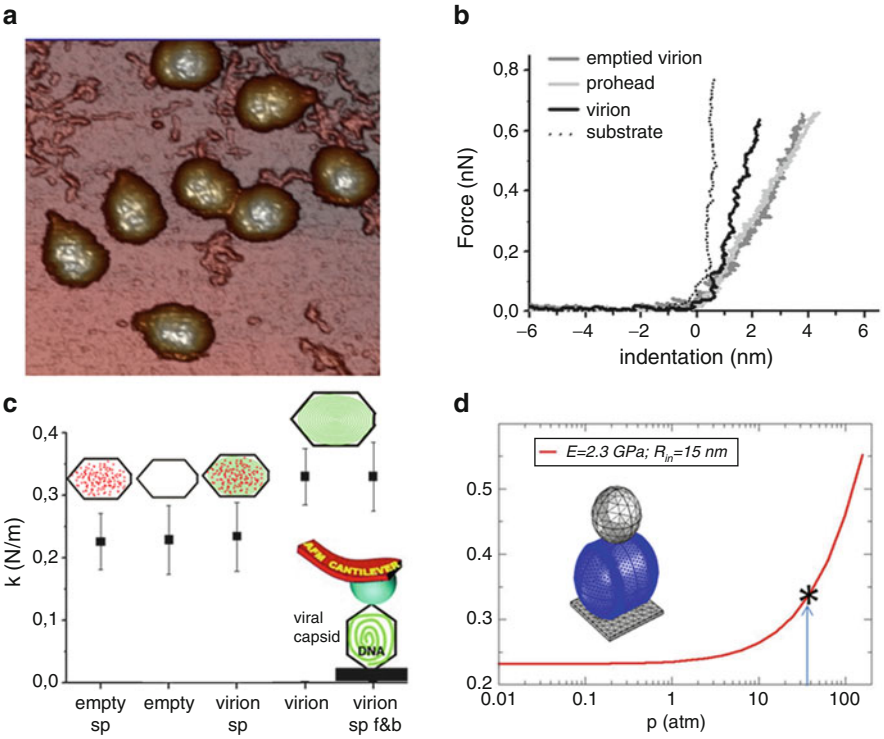


Fig. 18.4 Phage ϕ29 internal pressure. (**a**) AFM topography of mature ϕ29 particles; the tail is clearly visible. (**b**) Individual indentation curves on the substrate (*dotted*), DNA-filled ϕ29 virion (*black*), empty ϕ29 virion (*dark gray*) and ϕ29 prohead (*gray*). (**c**) Values of the spring constants obtained for the different types of ϕ29 particles analyzed in (**b**); sp, spermidine; sp f&b, spermidine added and later removed. (**d**) FEA results of the spring constant as a function of pressure for a phage ϕ29-like structure

Residual mechanical stress is the stress remaining after its cause has been removed. If we take a straight rod and simultaneously push from both ends (Fig. 18.5a), it bulges up in a stressed bump. This deformation quickly disappears when forces are removed, and the rod recovers the original straight shape. However, this stress can be conserved if both ends of the bulged rod are pinned.

The shell of bacteriophage ϕ29 prohead exhibits an intriguing anisotropic stiffness: Nanoindentations show that the prolate ϕ29 prohead is about twofold stiffer along the short axis (particle laying on its side) than along the long axis (upright particle) (Fig. 18.5b). The experimental and theoretical evidence indicates that this anosotropy is due to residual mechanical stress, as discussed next.

If the ϕ29 prohead is considered as a shell, continuum elasticity theory (see Sect. 18.3.1 and Chap. 19) predicts spring constant opposite trends along the short and long axes of the empty virus particle to those revealed by experiment. The ϕ29 capsid can be considered as a shell made by a cylindrical body closed by two

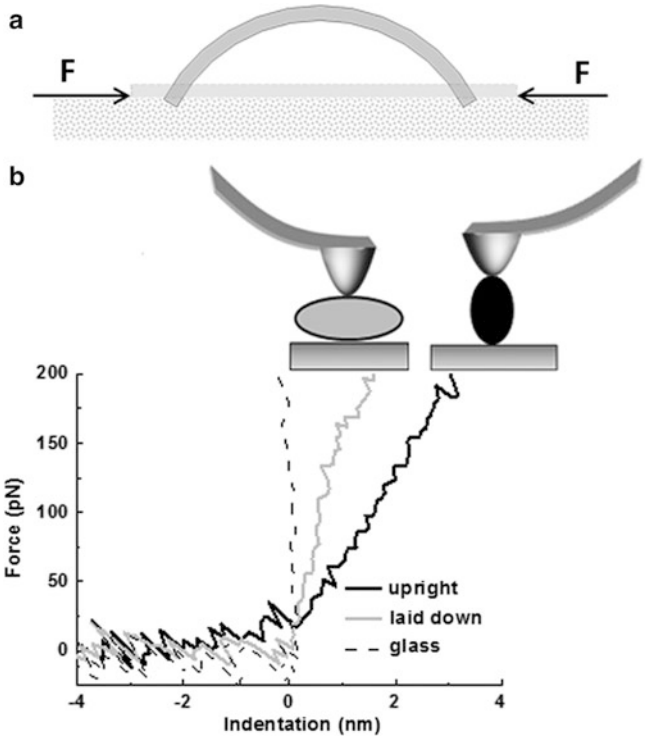


Fig. 18.5 Built-in mechanical stress. (**a**) The concept of residual stress. (**b**) Experimental results of indentations on ϕ29 particles laying on its side (*gray*) or upright (*dark*). Interpretation of these results indicate the presence of built-in mechanical stress in these particles (see text)

spherical caps. By considering independently the elastic response of the cylindrical and spherical parts of the virus particle one would expect that the laid-down virus would be mechanically softer than the upright one, since a cylindrical shell is easier to deform than a spherical shell of the same dimensions [18]. This is because a cylindrical shell, unlike a spherical one, can be bent without much stretching. We have already seen that, for a spherical shell, the effective spring constant is $k_{eff}^{sph} \sim Eh^2/R$. However, when indentation of a cylinder is considered [12], the spring constant equals $k_{eff}^{cyl} \sim Eh^{5/2}/R^{3/2}$. Therefore, the ratio of spring constants is $k_{eff}^{sph}/k_{eff}^{cyl} \sim \sqrt{R/h}$, which for the geometry of ϕ29 ($R = 21\ nm$ and $h = 1.6\ nm$ [35]) becomes approximately $k_{eff}^{sph}/k_{eff}^{cyl} \sim 3.6$. Hence, in the framework of continuum elasticity theory of shells, the cylindrical body of the virus is expected to be more than three times softer that the spherical caps. This clear disagreement with the experimental results can be attributed to the existence of residual stress on the equatorial zone of the shell imposed by the protein curvature, a hypothesis that was confirmed by coarse-grained simulations [31].

This built-in stress in the virus prohead could be an evolutionary strategy to provide extra mechanical strength. It is interesting to discuss prestress generation during the formation of the capsid. In the assembly process there is a competition between the tendency of proteins to aggregate at the preferred curvature and the need to minimize the rim area exposed in a partially assembled capsid by completing a closed structure [40]. If the curvature of the closed capsid is different from the spontaneous one, lateral stress will develop. Interestingly, not only ϕ29 [41], but other complex dsDNA bacteriophages have a scaffolding protein that co-assembles with the major capsid protein to produce a prohead with the correct shape and size [42] (see Chap. 11). Once the shell is built, the scaffolding proteins are released concomitant to DNA packaging. The possibility that some scaffolding protein subunits might be released from the proheads used in the experiments during previous storage cannot be excluded, but no change in prohead structure was in fact detected. Since protein-protein binding interactions in viral capsids are relatively weak, on the order of a few k_BT [43] (see Chap. 19), it is likely that scaffolding proteins help to impose a curvature in the capsid very different from the spontaneous one, by assisting the bending of the proteins at the junctions between subunits. This will generate a much larger in-plane stress, which will mechanically reinforce the capsid. In the absence of a scaffold, the stress generated during assembly could help to better tolerate DNA packaging, avoiding capsid disruption.

18.6 Differences and Variations in Virus Mechanical Properties

The studies on phage ϕ29 mechanics described in Sect. 18.5, and related studies carried out with a dozen other, structurally different model viruses so far (Fig. 18.6) [44] reveal that some fundamental mechanical concepts may apply to virus capsids in general. However, they also reveal that virus capsids may differ widely in specific mechanical features; as much as they may differ in structure, other properties, or the way they function during the viral life cycle. Comparison of the mechanical features of different viruses is helping in the development of theoretical models, and in the formulation of hypothesis on the structural foundations of the mechanical properties of viruses and their possible roles in virus biology. Such models and hypothesis are being subjected to experimental verification by using physical, chemical and/or molecular biology techniques including those described in this book (Chaps. 3, 4, 5, 6, 7, 8 and 9).

18.6.1 Differences in Stiffness

All viral capsids are made up of a same material, protein, and the capsids of about half the virus families can be roughly approximated to hollow spheres with a relatively thin crust made of protein tiles. Yet, determination of the spring constant k of the

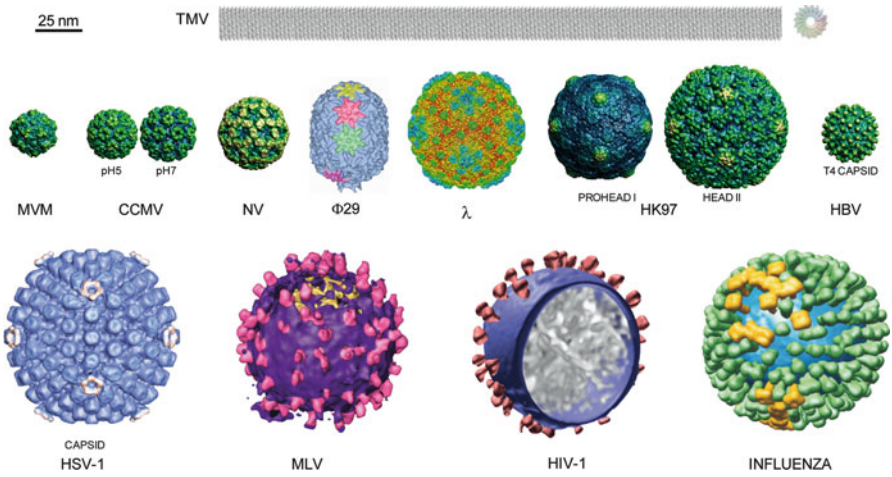


Fig. 18.6 Virus particles whose mechanical properties have been studied to date (see Sects. 18.5 and 18.6). All particles are reproduced at (approximately) the same scale, indicated by the bar at *top left*. Side and top views of TMV, native and swollen forms of CCMV, and prohead-I and head-II forms of HK97 are represented. Fibers in ϕ29 and tails in ϕ29, λ and HK97 virions are not shown. For HBV (T = 4 form) and HSV-1, mechanical properties were analyzed for naked capsids only, and these (not the enveloped virions) are represented. This figure is reproduced from [44] with permission from the publisher. The different viral particle images were originally taken from VIPERdb [45] or from the references indicated in the original figure in [44], and are reproduced with permission from the publishers

dozen or so spherical virus capsids experimentally studied to date using AFM reveals that their mechanical stiffness can differ by close to two orders of magnitude: from 0.018 N/m for the prohead-I of bacteriophage HK97 (as low as that of a "soapy" lipidic viral envelope) to as much as 1.3 N/m for some minute virus of mice (MVM) capsid mutants. How could we explain such large differences?

A quick look at the molecular structures of spherical viruses determined by cryo-EM (Chap. 3) or X-ray crystallography (Chap. 4) reveals that virus capsids are very different from each other in overall size, thickness and the number, size, shape and geometric arrangement of the protein tiles they are made of (Chap. 2 and Fig. 18.6). A thicker hollow sphere will be more difficult to deform by pushing with the tip of the AFM than a thinner one made of the same material. A smaller hollow sphere will be more difficult to deform than a larger one having the same thickness and composition. The measured mechanical stiffness of a virus capsid, or any other solid object, as given by the spring constant, depends not only on the material it is made of, but also on its dimensions and geometry (in the widest sense of the term).

18.6.2 Differences in Intrinsic Elasticity

It may be more useful for our discussion to consider not the spring constant of the capsid, but the *intrinsic* elasticity of the capsid material, the protein crust, as given

by its Young's modulus (see Sect. 18.3 and Chap. 19). Unfortunately, as already discussed, some simplifications are needed to calculate the Young's modulus of a virus capsid. Although many icosahedral virus capsids can be approximated to thin-shelled perfect spheres or icosahedrons, an inspection of the capsid structure reveals that virus capsids are ridden with protrusions and depressions, sometimes very conspicuous. Capsid thickness is not uniform, either. However, for some virus capsids, the approaches described in Sect. 18.3 or other approaches yielded reasonably similar Young's modulus values.

The average Young's moduli obtained for different spherical virus capsids still span over one order of magnitude, from 0.14 GPa for the cowpea chlorotic mottle virus (CCMV) capsid to 1.8 GPa for the phage Φ29 prohead (see Sect. 18.5.1), and up to 2.8 GPa for some MVM capsid mutants. From these values, some of the hollow quasi-spherical protein shells of viruses could mechanically resemble soft rubber balls, while others could be compared to hard acrylic plastic globes. The actual situation is even more heterogeneous. Some viral capsids (those of phage Φ29 or some MVM mutants) are mechanically anisotropic, and discrete regions in them may show up to severalfold differences in spring constant [31, 46]. Thus, a single value for the Young's modulus may not be appropriate to describe the capsid intrinsic elasticity. In addition, for a same viral capsid in different conformational states (phage HK97; [47]) or carrying different single mutations per subunit (MVM; [48]), the Young's modulus may vary up to severalfold. Thus, the protein material the nanoscopic viral capsids are made of, contrary to the material that constitutes many macroscopic solid-state objects, is able to substantially change its intrinsic elastic properties through very minor modifications in composition (*e.g.*, a single amino acid substitution per subunit) or the spatial rearrangement of (some of) its molecular components.

18.6.3 Differences in Brittleness and Resistance to Material Fatigue

As for mechanical stiffnesss, spherical capsids differ widely in their ability to withstand high mechanical loads (forces) without being physically disrupted, or even irreversibly deformed. In general, mechanical failure (breakage) of a virus capsid appears to occur when it is deformed to a certain extent [36, 49]. Some capsids (phage Φ29 prohead, mature phage λ capsid, phage HK97 prohead-II, the MVM capsid, or the Norwalk virus (Norovirus, NV) capsid at basic pH; Fig. 18.6) were fractured or disrupted when the tip of the AFM was used to apply forces that caused not very deep deformations of the protein shell, generally not above 10–15 % of the capsid diameter. In contrast, other capsids (phage HK97 prohead-I, hepatitis B virus (HBV), NV at neutral or acidic pH, CCMV at pH = 6; Fig. 18.6) withstood forces that elicited very large deformations of the particle. Some of the latter withstood even a wall-to-wall collapse under load; once the force was

removed and the particles were allowed to relax, they generally recovered their original dimensions, with no structural damage being detected. Thus, some viral protein shells can be regarded as relatively brittle, while others are considerably more resilient under mechanical loads. Differences between viral capsids regarding material fatigue have also been noted. Like stiffness, brittleness and resistance to material fatigue of a viral capsid have been observed to depend on the conformational state of the particle. However, it must be stressed that systematic, controlled and precise estimations of the brittleness or resistance to material fatigue of virus particles are still lacking.

Stiffness, brittleness and resistance to material fatigue in viral capsids do not appear to be correlated, and may also vary in complex ways (*e.g.*, as in phage HK97; [47]). A same capsid may be quite stiff but, at the same time, relatively brittle and fail under relatively low mechanical loads causing small deformations. Another capsid may substantially change its brittleness without any variation in stiffness, and so on and so forth. This aspect also requires further investigation.

18.6.4 A Relationship Between Molecular Structure and Mechanical Properties of a Virus Capsid

As viral capsids are all made of protein, what structural differences cause the substantial differences in intrinsic elasticity (Young's modulus), brittleness and resistance to fatigue among them, and even within them? The logical interpretation advanced by several researchers is that neither capsid protein subunits, nor the lattice formed by these subunits in the viral capsid, are really a continuous, mechanically homogeneous material. The intrinsic elasticity of a viral capsid may depend on the number, distribution, directionality, type and energy of the covalent and non-covalent interactions between atomic groups and amino acid residues within a same capsid protein subunit, or between neighboring subunits. Covalent interactions include the linkages between chemical groups and residues in each polypeptide chain and, in a few capsids, disulfide bonds or other covalent crosslinks between residues. Non-covalent interactions are extremely numerous and include very different types, such as van der Waals contacts, hydrogen bonds, ionic interactions, and the entropically driven hydrophobic effect. The above molecular interpretation has been invoked to explain, for example, the pH-dependent reduction in stiffness of the CCMV capsid [21, 50]; the large increase in stiffness of the phage HK97 capsid during its initial maturation stage (prohead-I to prohead-II) [47]; or the remarkable (up to 120 %) increases in mechanical stiffness caused by a single amino acid substitution (per protein subunit) in the MVM capsid [48].

In summary, virus capsids can vary widely in intrinsic stiffness, structural strength under load and/or, probably, resistance to material fatigue. Such mechanical differences may be due to either large, or sometimes small differences in molecular structure, including number, type, strength and distribution of inter- and intrasubunit covalent and non-covalent interactions.

What specific structural differences are responsible for the observed mechanical differences between two viral capsids? If we can answer this question at the molecular (atomic) level (Sect. 18.7), the roles of the mechanical features of a viral capsid during the virus life cycle could be explored and elucidated (Sect. 18.8). It could also facilitate the rational modification of the mechanical properties of viral capsids and other protein nanoparticles for biotechnological purposes (Sect. 18.9).

18.7 Structural Determinants of the Mechanical Properties of Viruses

Despite the complex relationship between molecular structure and mechanical properties of virus capsids, theoretical and experimental approaches, and combinations of them, are starting to unveil some aspects of that relationship. However, quite a lot remains to be done in this novel area of research. Yet, the few studies to date have already provided some facts and many testable hypothesis that hold the promise of a better understanding of virus particles from the points of view of fundamental physics and structural biology.

18.7.1 Structural Determinants of Capsid Mechanics: A Theoretical Approach

A very considerable number of physics-based theoretical analyses of virus mechanics have been undertaken in the last years to propose fundamental physical explanations of different experimentally observed mechanical features of virus particles, including some of those described in this chapter (see [49, 51, 52] and Chap. 19). In addition, theoretical analyses and modeling are leading to predictions related to virus mechanical properties and their materials foundations [53], most of which have not been experimentally addressed yet. Many such theoretical analyses are based on elasticity theory using simplified models, finite element or coarse-grained molecular dynamics (MD) simulations, and are described in detail in Chap. 19.

18.7.2 Structural Determinants of Capsid Mechanics: An Experimental Approach

Structural Variations and Changes in Mechanical Behavior

An experimental structure-function relationship approach has been followed by several research groups to investigate, at different levels of resolution, the structural

foundations of mechanical features in viral capsids. In this type of studies, a difference in mechanical behavior may be traced to a structural difference between two similar capsids.

So far, most of the yet scarce virus structure-mechanics comparisons of viral particles have been made between structurally or conformationally quite different particles. For example, between different global conformational states of the CCMV and HK97 capsids [21, 47, 50]; or between the NV or the herpes simplex virus type 1 (HSV-1) capsids with or without a large protein part removed [36, 54]. In such cases, the large structural differences make it very difficult to identify with high resolution the specific molecular foundation(s) of the mechanical difference considered.

Ideally, the higher the structural similarity between the capsids compared, the better the chance to find the true molecular cause of any change in a physical property or function. Either two evolutionarily very closely related capsids can be compared or, better yet, rationally chosen single amino acid substitutions can be introduced by site-directed mutagenesis in a given viral capsid, and their effect on capsid structure and mechanics analyzed. Unfortunately, only in very few cases this high-resolution structure-function approach has been undertaken so far to study the molecular basis of virus mechanics.

In a study on CCMV, a single amino acid substitution (K42R) at the N-terminal arm of the capsid protein was shown to introduce 660 new intersubunit interactions in the whole capsid; the new interactions were invoked to explain both the higher mechanical stiffness compared to the non-mutated capsid, and the increase in stability against capsid dissociation when high salt concentrations are present [55].

In another study, the individual replacement of several amino acid substitutions in the MVM capsid by alanine, which involved removal of a single chemical group per subunit, led to a drastic stiffening of the viral capsid [48]. The group removed was involved in very few noncovalent interactions within each capsid subunit, or between pairs of subunits, and the mechanical effect was tentatively attributed to a conformational rearrangement in a metastable capsid.

In a third study, the decreased mechanical stiffness of the NV capsid at acidic pH was attributed to deprotonation of a specific lysine residue that could introduce an electrostatic repulsion between certain aspartate residues, triggering a conformational rearrangement [56]. This prediction could be tested by site-directed mutagenesis of the residues putatively involved in the pH-dependent effect on mechanics.

Additional experimental structure-function studies, as well as all-atom MD simulations of whole viral capsids once the latter becomes more feasible (see Chap. 19), may help to ascertain at high resolution the structural determinants of the mechanical behavior of viral capsids, and provide insights into the relationship between mechanical, thermal and chemical stability in a capsid or any other nanoparticle.

Molecular Interactions and Capsid Fracture

As described in previous paragraphs, virus capsids can be mechanically irreversibly modified, even fractured, under a high enough and/or repeated load exerted, for example, by the AFM tip (Fig. 18.7). Such fractures are observed in

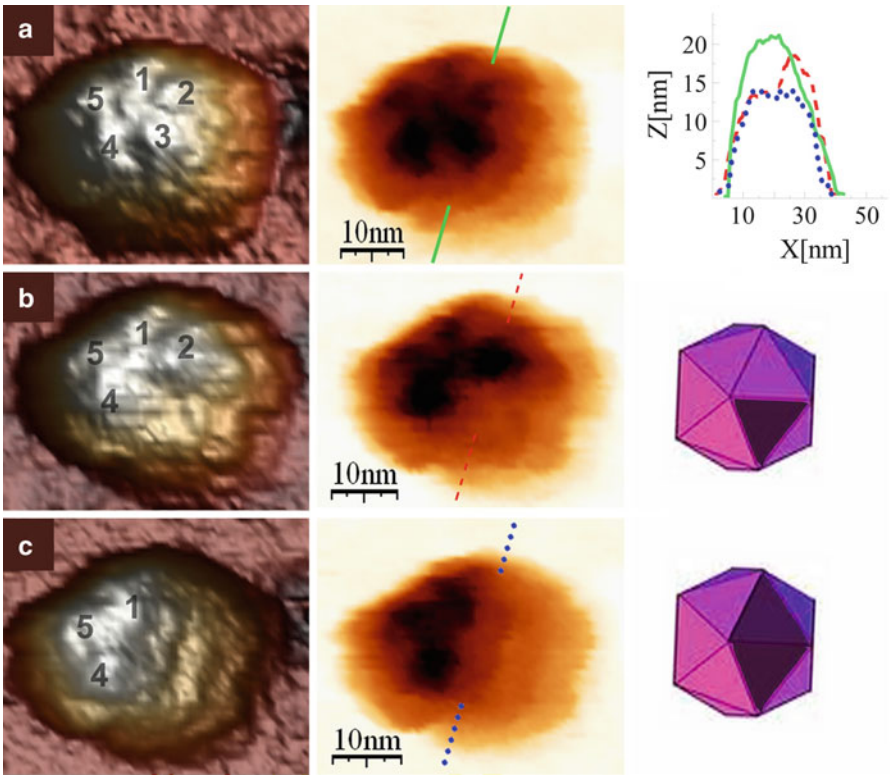


Fig. 18.7 Mechanical disassembly of a simple virus. AFM images of a same individual MVM capsid before indentation (**a**), after removal of one capsid building block (a trimer of capsid proteins) (**b**) and after removal of an additional trimer (**c**) *Left*, 3D images; *center*, negative images. *Top right*, height profiles of the intact capsid (*green line*), of the capsid after losing the first trimer (*red dashed line*), and of the capsid after losing an additional trimer (*blue dotted line*). *Bottom right*, schemes of the MVM capsid after losing one trimer (*top image*) or two trimers (*bottom image*) (Reproduced from [57]. With permission of the publisher)

nanoindentation experiments as nonlinearities in the FZ curves obtained during the indentation. In some cases (*e.g.*, NV capsid) these fractures may occur but the capsid recovers ("self-heal") after the load is removed [56]. In other cases, more or less permanent fractures may occur. Visualization of these fractures in different viral capsids [Φ29, HSV-1, MVM (Fig. 18.7), adenovirus] by AFM indicates that they may occur along the borders of capsid subunits or assemblies of capsid subunits that may act as intermediates of capsid assembly and/or disassembly [24, 36, 57]. In some cases, the dissipated energy during those fractures was estimated and found roughly comparable to the estimated association energy between capsid subunits [58].

The results obtained so far are scarce, but they suggest that mechanical failure (fracture) in virus capsids could generally start by the disruption of many (or most) noncovalent interactions at the energetically weakest interfaces between capsid subunits or assembly/disassembly intermediates, leaving the energetically more stable

building blocks intact. The intersubunit interactions thus disrupted may re-form relatively fast, leading to self-healing. Alternatively, further interfaces may be disrupted. If all of the interfaces between a single capsid building block and its neighboring blocks are eventually disrupted, the building block will be released (as observed with MVM (Fig. 18.7) and adenovirus). If additional interfaces are disrupted by a continuous or repeated high load, the capsid will collapse, also as repeatedly observed.

Loss of capsid subunits (Fig. 18.7) and self-healing of viral capsids by restoration of disrupted interfaces, have been predicted by thermodynamic-kinetic models [59] and coarse-grained MD simulations [60] (see Chap. 19) of reversible capsid (dis) assembly in solvent. According to these models and simulations, capsids that lose one capsomer under mechanical load or chemical intervention tend to reassociate with the released capsomers, acting as kinetic traps that prevent further dissociation of the capsid. Thus, thermodynamically unstable capsids that would dissociate at low particle concentrations, as found in the extracellular stage of the virus life cycle) are stabilized, and can preserve its physical integrity ([40, 43] and Chap. 19).

18.7.3 Modulation of Virus Particle Mechanics by Components Different from the Capsid

In the preceding paragraphs, only the mechanical properties of the protein capsids of viruses have been considered. What about the effect of other viral particle components on the mechanical behavior of the complete virion?

The Viral Nucleic Acid

As described above, the mechanical behavior of some viral capsids can be substantially modified if they contain a viral nucleic acid packaged inside, as in complete virions. In MVM, CCMV, and phages λ and ϕ29 (Sect. 18.5.2) the nucleic acid stiffens the capsid [30, 34, 46, 55, 61, 62], but in HSV-1 no mechanical stiffening was detected [36]. In many spherical virions, the hydrated nucleic acid occupies a very large fraction of the capsid interior, not infrequently reaching crystalline densities. Under such conditions, the molecular basis for the observed nucleic-acid mediated increase in rigidity of several virus capsids could, in principle, include steric effects, electrostatic repulsions between the phosphates, the high resistance to bending (given by the persistence length) of the nucleic acid molecule if double-stranded, hydration forces, etc. (see Chap. 19). Interestingly, the studies carried out reveal that the molecular determinants of the nucleic-acid mediated mechanical stiffening effects in different viruses can be quite different.

The high internal pressure (about 40–60 atm) exerted by the tightly packaged dsDNA inside phage Φ29 contributes to increase the stiffness of the particle, not unlike the stiffness of a football is increased by the air compressed inside ([34], see Sect. 18.5.2 for a description). Wild-type phage λ also showed a dsDNA-mediated increased stiffness. However, two mutant λ virions that contained somewhat

shortened dsDNA molecules, which would still exert a substantial internal pressure, showed no higher stiffness than the empty particle [52, 61]. Based on their results and calculations, Ivanovska et al. [61] proposed that, because of DNA-hydrating water molecules, an osmotic pressure exists in the λ capsid that increases rapidly when the DNA density approaches that in the wild-type phage, and it is this osmotic pressure what increases the stiffness of the particle.

The single-stranded (ss) DNA in MVM is also packaged in the capsid (Chap. 10) to densities that approach those found in dsDNA phages. However, the drastic ssDNA-mediated stiffening of MVM [30] is due not to internal pressure or hydration forces, but to a buttressing effect of short DNA segments that bind a number of equivalent sites at the capsid inner wall [46].

Unlike the above viruses, the ssRNA of CCMV is not packaged in a preformed capsid, but both are assembled together (Chap. 12), and the CCMV virion may not be pressurized. However, the ssRNA inside does increase the mechanical stiffness of the CCMV particle [55] by a yet unknown mechanism.

The Viral Envelope

Some mechanical features of three complete enveloped virions have been investigated. Again, remarkable differences were encountered. The influenza virion (Fig. 18.6) is extremely elastic, mechanically behaving almost like its very soft lipidic envelope [63, 64]. In contrast, the immature virions of two retroviruses, murine leukemia virus (MLV) and human immunodeficiency virus (HIV-1) (Fig. 18.6), are mechanically very stiff [65, 66], yielding estimated Young's moduli comparable to those of icosahedral capsids.

The available evidence suggests that the large stiffness differences between those viruses are not due to the lipid envelope, but depend on the structural organization of the protein layers contained within. In the influenza virion, the matrix M1 protein layer immediately below the envelope may be highly flexible and will not oppose any significant resistance to deformation. In contrast, the very thick Gag capsid below the envelope of retroviruses would be partly responsible for the high spring constants determined for the immature MLV and HIV-1 virions. In addition, the even higher stiffness of the immature HIV-1 compared to MLV has been traced to the Env protein embedded in the envelope, perhaps due to the presence of Env-Gag interactions that may not occur in the immature MLV, although this possibility remains to be further investigated.

The remarkable decrease observed in MLV and HIV-1 stiffness during virion maturation has been attributed, at least in part, to the disassembly of the thick immature Gag capsid, leaving a much thinner matrix (MA) protein layer under the envelope. The reassembled CA protein capsid has a conical shape and is much smaller than the virion; thus, the latter could be mechanically deformed without reaching contact with the loose, smaller capsid contained inside. In the HIV-1 case, maturation would additionally disrupt the proposed Env-Gag interaction, contributing to the extreme mechanical softening observed in this case [65, 66]. It may be noticed here that the three infectious enveloped viruses analyzed are mechanically very flexible, and this may not be a coincidence (see Sect. 18.8).

18.8 Mechanical Properties and Virus Biology

The solid-state physicist and the structural biologist/chemist may appreciate the value of the emerging results on the mechanical properties of viruses and their structural determinants for a better understanding of "soft" condensed matter and self-organizing molecular assemblies. However, the virologist may question whether the specific mechanical characteristics of virus particles are merely a consequence of their being solid objects, or are shaped by evolution because they confer some selective advantage to the virus. This question is still open, but growing evidence indicates that virus mechanical features may be biologically critical in more than one way.

If the mechanical features of viruses have been shaped by evolution, what are the selective pressures behind such evolution? Are viruses subjected to mechanical stress that may select for mechanically more robust variants? Do virus particles require a certain mechanical elasticity to complete some step in their life cycle? Indeed, virions are subjected to physical forces that may deform or break them either outside or inside cells. Such forces include shear forces while drifting in viscous fluids, or being extruded through nuclear pores or other openings; osmotic effects; pressurization due to packaged dsDNA; hydrostatic pressure; capillary forces on desiccation, etc. Viruses with a tubular, very elongated shape, such as tobacco mosaic virus (Fig. 18.6) may be particularly sensitive to mechanical stress, but any virus, because it is made of "soft" material, may seriously risk mechanical disruption at some point or other in its existence. Moreover, some virus particles must change its shape (undergo conformational rearrangements) to fulfill its biological function (see Chaps. 1, 13 and 15); thus, their ability to deform or withstand deformation, as determined in mechanical measurements, may be connected with their biologically relevant conformational dynamics.

We have stressed in previous sections the remarkable differences in mechanical features when different virus particles, or a same particle in different stages of the infection cycle, are compared. It seems reasonable to propose that viruses may have evolved different mechanical features in response to the selection pressures of the different forces acting on them, to withstand those forces, or even for using them to their own advantage. In this section we provide three case studies that strongly support different biological roles for different mechanical features found in virus particles.

18.8.1 Mechanical Stiffening to Withstand Pressurization

At least in some dsDNA viruses, including phages Φ29, λ and HK97, and the herpesvirus HSV-1 (Fig. 18.6), the genome is actively packaged inside a preformed capsid that is barely large enough to contain the hydrated full-length nucleic acid molecule. As a consequence of the very high packing density, the dsDNA inside exerts a very high internal pressure that tends to disrupt the capsid. Several authors

have noticed that the intrinsic stiffness of these viruses is relatively high (Young's modulus 1.0 GPa or higher). In contrast, in several (albeit not all) viruses that enclose their viral ssRNA (much less stiff than dsDNA) in a condensation process (Chap. 12), such as HBV, CCMV, influenza virus, and immature MLV, the Young's modulus is low (0.37 GPa or lower). One could ask why the pressurized dsDNA virus have not evolved to decrease their high internal pressure by slightly shortening its DNA molecule. In fact, pressurization in the dsDNA tailed phages (at least) has a biological function by driving injection of the viral DNA into the host bacterium [52, 67]; see Sect. 18.5.2 and Chaps. 17 and 19. Pressurization in HSV-1 may be likewise required for the translocation of its DNA into the cell nucleus through a nuclear pore [67].

The available evidence has led to the proposal that dsDNA phages and other viruses may have evolved a capsid of relatively high intrinsic stiffness as a biological adaptation to withstand the high internal pressure required for DNA injection [31, 49, 52, 61, 67]. During maturation (Chap. 13), the capsid of these phages expands and, as a consequence, is made thinner. Instead of making the capsid thicker, these phages may have solved the problem of mechanical weakening during maturation by making the capsid of a particularly strong material. This high intrinsic mechanical resistance may be due to strong interactions between capsid subunits. In addition and in agreement with this view, it has been observed that, during maturation, expansion and thinning is accompanied by the introduction of external reinforcements, like a chainmail of protein crosslinks (HK97) or cementing proteins (λ) [47], or even by prestressing the capsid (Φ29) [31]. The results suggest that successful maturation of at least some dsDNA viruses may critically depend on the mechanical features built-in in their capsids, and on the appropriate variations in mechanical resistance through the controlled introduction of certain structural modifications.

18.8.2 Mechanical Softening to Allow Entry into Cells

Immature MLV and HIV-1 virions enter host cells very inefficiently compared with the mature, infectious virions. Interestingly, maturation makes these viruses much less rigid (as determined by their spring constant) [65, 66]. In the case of HIV-1, it was also observed that a specific structural modification introduced in the laboratory, the truncation of the carboxy-terminus of the Env protein promoted the efficient entry of immature HIV-1 virions into cells and reduced the extreme stiffness of the immature virion to a value that approached that of the mature virion [66]. Based on those results, it has been hypothesized that effective fusion with the cell membrane to allow virus entry (see Chap. 16) may require a mechanically soft, flexible enough virion for extensive virus-cell contact. The structural changes effected during virus maturation, specifically a probable loss of Env-Gag interactions, would have as one consequence the mechanical softening of the virion, that would be required for virus entry. The high mechanical elasticity of a very different enveloped virus, influenza virus, is also consistent with the above

hypothesis. If this hypothesis is verified, changes in mechanical features of enveloped virus could effect a mechanical control of entry of enveloped viruses into host cells.

NV, a nonenveloped virus, must withstand the acidic conditions encountered during passage through the stomach, but must be able to dissociate to release its genome at weakly basic conditions during infection within the ileum. Remarkably, the NV capsid was found to be stiffer at neutral and acidic pH and less rigid at alkaline pH [56]. The authors suggested that the pH-dependent mechanical softening of the NV capsid could result from a putative, pH-triggered conformational rearrangement needed for infection, or even constitute a direct requirement to facilitate virus entry and RNA release. If this hypothesis is correct, mechanical features could have a biologically critical role also during the entry of some nonenveloped viruses into host cells (see Chap. 15).

18.8.3 Balancing Virus Stiffness to Prevent Inactivation Without Impairing Infection

In the MVM virion, segments of the viral ssDNA bound to equivalent sites at the capsid inner wall act like molecular buttresses that increase the mechanical stiffness of most regions in the virus particle. However, the regions around pores located at the fivefold symmetry (S5) axes of the icosahedral capsid are free from bound DNA, and remain mechanically as flexible (soft) in the DNA-filled virion as in the empty (DNA-free) capsid [30, 46]. Is this anisotropic distribution of mechanical stiffness in the virion due to mere chance, or has it evolved because it provides some biological advantage? The available evidence (summarized next) supports a biological role for the anisotropic stiffness of the MVM virion.

Previous studies had revealed that the capsid-bound DNA segments do provide the MVM virion with the biological advantage of increased resistance against thermal inactivation, which appears to be mediated by a conformational change in the virion; it was suggested that the interactions between the capsid and DNA segments may impair this conformational change [68]. In a series of different studies it was found that the MVM capsid regions around the S5 pores are structurally dynamic, as they participate in biologically relevant conformational rearrangements. These involve opening and closing of the capsid pores for translocation of peptide segments carrying molecular signals, and also the encapsidation (Chap. 10) and uncoating (Chap. 15) of the viral DNA. Mutations of residues at the base of the pores in the MVM capsid impair the biologically relevant conformational dynamics of the S5 (pore) regions and dramatically reduce virus infectivity at physiological temperature [69]; likewise, mutations of residues lining the S5 pores of adeno-associated virus type 2, a related virus, also impair local dynamics.

Those observations led to the hypothesis that in a virus capsid, high mechanical flexibility, or softness (as determined by AFM) and high conformational dynamism

(identified by other techniques like limited proteolysis/mass spectrometry or intrinsic fluorescence, see Chap. 6) are two descriptions of a same molecular feature, *structural flexibility*: If a region of a viral particle can be varied from its minimal free energy conformation without requiring too much energy, a structural biologist may detect a conformationally dynamic region prone to undergo some structural rearrangement (see Chap. 1); while a condensed matter physicist using AFM may detect a mechanically soft region with a relatively low spring constant. The phenomenon would be the same, but it would be detected in different ways by probing different properties. If this hypothesis is correct, regions of higher mechanical flexibility (softness) in a virus particle could match conformationally more dynamic regions.

Recently, the above prediction was verified by independently testing the effects of many single amino acid substitutions on capsid stiffness, the impairment of specific conformational rearrangements (as a measure of capsid dynamics) and virus infectivity. As predicted, mutations at the base of the capsid pores that impaired the local conformational dynamics invariably increased the mechanical stiffness of the regions around the pores [48]. Also, the two mutations tested so far that facilitated the heat-induced non-productive conformational rearrangement of the virion decreased the stiffness of most regions in the virion (but not the regions around the pores), nearly abolishing the DNA-mediated rigidification of the virion [46]. Mutations that had no effect on conformational dynamics had no effect on stiffness, and further mutations that restored a lost conformational transition also restored the lost mechanical softness [48], again as predicted by the hypothesis that high conformational dynamism and high mechanical flexibility in virus particles may be phenomenologically linked.

The above results provide strong experimental support for a biological role of the anisotropic distribution of mechanical elasticity in the MVM virion. MVM may have evolved DNA-binding sites in its capsid because the DNA-mediated structural rigidification of most capsid regions increases virion thermostability (by impairing a virion-inactivating conformational rearrangement). In contrast, the regions around the capsid pores may have been kept free of bound DNA by negative selection, because a rigidification of these regions would impair the local S5 dynamics required for infectivity. In short, the anisotropic mechanical stiffness in the MVM virion reflects the unequal distribution of structural flexibility/conformational dynamism in the viral particle, and may constitute a biological adaptation that impairs virion inactivation without impairing infection.

18.9 Engineering Mechanical Properties of Virus Particles

There is a rapidly growing interest in the use of viral particles for different biotechnological or nanotechnological applications (see Chaps. 21 and 22; [70]). However, it is envisaged that, for quite a few of such applications, natural viral particles may be not physically resistant enough. Potentially useful nanoparticles may be subjected to high

mechanical stress during purification, fabrication, storage and/or use. Such conditions are very demanding for "soft" biological materials and may lead to the disruption of natural viral particles. In addition, many natural viral particles are metastable and particularly sensitive to conformational rearrangements (see Chaps. 1 and 13; [71]). Conformational instability of a virus particle may be desirable for some bio/nanotechnological applications. For example, virus particles for gene therapy or targeted drug delivery may have to be capable of dynamic rearrangements required to deliver their cargo into cells. Even in those cases, improved mechanical stability may be highly desirable to withstand physical stresses during production and purification. On the other hand, structural rearrangements may lead to functional inactivation. For many potential applications including contrast agents, diagnostic agents, vaccines, building blocks for nanomaterials, etc., viral particles may not need to be conformationally dynamic. In such cases, the highest possible mechanical stiffness to withstand permanent deformations, and the highest structural strength against mechanical disruption, as well as thermal and chemical stability, may be critical.

The physical, chemical and/or genetic manipulation of virus particles, generally carried out to investigate the structural basis or the biological relevance of specific mechanical features, have led as a by-product to the mechanical destabilization of the manipulated capsids. More elastic and/or more brittle particles have been obtained by truncating a several-domain protein [54, 66]; eliminating some capsomers [36, 57]; eliminating or reducing the length of the nucleic acid molecule [61]; neutralizing the nucleic acid negative charge [62]; or eliminating capsid-nucleic acid interactions [46].

However, mechanical stabilization may be much more useful than destabilization when considering viral particles for most nanotechnological applications. Research on the structural elements that underlie the mechanical features of natural viral particles may provide useful guidelines for engineering mechanically more robust viral nanoparticles. For example, it was found that during maturation of phage λ, protein gpD attaches to the capsid, acting as a cement that stabilizes the particle [72]. In phage HK97, a chainmail of covalent crosslinks between capsid subunits contribute to increase the structural strength and resistance to material fatigue and physically stabilize the particle [47]. In CCMV, a single amino acid substitution biologically fixed in a variant of CCMV conferred resistance of the capsid to high salt concentrations and also increased its stiffness [55].

A rational approach based on biophysical and biological evidence has been successfully used to increase the mechanical stiffness of a viral capsid by protein engineering. Some amino acid substitutions in the MVM capsid that impaired conformational dynamics were predicted to increase mechanical stiffness (see Sect. 18.8.3). In addition, substitutions of amino acids surrounding conspicuous cavities in the MVM capsid were also designed to modify the size and shape of such cavities, and some of them were predicted to increase capsid stiffness. When tested, many of these mutations did increase, as predicted, the capsid mechanical stiffness (up to close to threefold higher spring constant and Young's modulus, compared with the natural empty capsid). Some of these engineered mutant capsids were, by far, stiffer (1.3 N/m, 2.8 GPa) than any other naked capsid analyzed to date [48]. These studies provide the only examples so far of the purposeful mechanical stabilization of

a viral particle by protein engineering in the laboratory. They constitute proof-of-principle that viral nanoparticles can be manipulated by protein engineering to adequately modify their mechanical characteristics, with the aim of improving their usefulness in bio/nanotechnology.

18.10 Perspectives and Conclusions

Fifteen years have passed since a pioneering study on virus mechanics in air using the rod-like tobacco mosaic virus as a model was published [73]; the first study on the mechanical properties of small spherical viruses in liquid appeared only 8 years ago [6]. In this short time, several outstanding mechanical features of virus particles have been discovered, including the high elasticity, resilience and resistance to material fatigue of some (albeit not all) virus capsids; the mechanical anisotropy of virus particles; the nucleic acid-mediated mechanical buttressing in some virions, and the different structural foundations of such effects; the drastic effects on capsid elasticity of single amino acid substitutions, capsid subunits or components; and the existence of built-in stress.

The results obtained on virus mechanics are relevant in materials science, structural and physical virology and bio/nanotechnology. From the point of view of materials science, "soft" biological materials such as protein layers are being mechanically characterized. However, physical interpretations of some results may be not without caveats. Mechanical characterization includes the estimation of some volumetric physical magnitudes such as the Young's modulus, based on homogenous models of continuum mechanics. However, does it make any sense to define a Young's modulus for proteins based on a continuum approach? The relative Young's modulus can be useful, for example as a parameter for mechanical comparisons, but proteins are not a homogeneous material, and virus capsids themselves have very complex shapes, and are made of discrete subunits connected by different intersubunit interactions. Can mechanical experiment extract information about capsid subunits and their reciprocal interactions? Further experiments are required to gain insights into the relationship between the discontinuous nature of virus capsids and their mechanical properties. Mechanical parameters extracted from the experiments, such as the spring constant, are appropriate for mechanical comparisons between virus particle mutants, different regions in a same capsid, capsids in different maturation stages, and empty capsids *vs.* nucleic acid-filled virions. Fatigue material in virus particles has been little explored so far; experiments aimed at the mechanical disassembly of viral particles may provide information in this respect. Controlled mechanical dissociation of virus particles may also reveal assembly/disassembly intermediates; however, these experiments should be carefully controlled to allow the characterization of the intermediates thus produced.

From the point of view of virus biology, growing experimental evidence is providing support for a functional relevance of the mechanical properties of viruses. Mechanical stiffening achieved through the establishment of capsid-nucleic acid interactions, built-in stress, specific noncovalent interactions or covalent crosslinks,

for example, may help virus particles to withstand mechanical forces acting on them in the extracellular environment or, in the case of dsDNA phages, as a consequence of the internal pressure exerted by the packaged nucleic acid. Changes in mechanical properties of retroviral virions during maturation could help entry into cells by facilitating virus-cell membrane fusion. The DNA-mediated anisotropic mechanical stiffness of the MVM virion reflects the unequal distribution of structural flexibility and conformational dynamism in the virus particle, which may have arisen as a biological adaptation to better resist inactivation without impairing infectivity.

From the point of view of bio/nanotechnology, studies on the molecular determinants of the mechanical stability of viral particles at single amino acid-resolution may provide insights for the engineering of virus-derived nanoparticles with increased physical resistance to undesirable structural deformations or disruption.

Acknowledgements We gratefully acknowledge J.L. Carrascosa, J. Gómez-Herrero, D. Reguera, C. San Martín and N. Verdaguer for collaboration on virus mechanics, advice and fruitful discussions; C. Carrasco, P.J.P. Carrillo, M. Castellanos, M. Hernando-Pérez, A. Llauró, A. Ortega-Esteban and R. Pérez for excellent work on virus mechanics; M.A. Fuertes and A. Rodríguez-Huete for great experimental assistance. M.A.F. kindly provided Fig. 18.6. Work in P.J.P.'s laboratory is funded by the Spanish Government (grants PIB2010US-00233 and FIS2011-29493). Work in M.G.M.'s laboratory is funded by the Spanish Government (grants BIO2009-10092 and BIO2012-37649) and Comunidad de Madrid (S-2009/MAT/1467), and by an institutional grant from Fundación Ramón Areces. M.G.M. is an associate member of the Institute for Biocomputation and Physics of Complex Systems, Zaragoza, Spain.

References and Further Reading

1. Shiroguchi K, Kinosita K (2007) Myosin V walks by lever action and brownian motion. Science 316:1208–1212
2. Moreno-Herrero F, de Jager M, Dekker NH, Kanaar R, Wyman C, Dekker C (2005) Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. Nature 437:440–443
3. Moody MF (1999) Geometry of phage head construction. J Mol Biol 293:401–433
4. Cordova A, Deserno M, Gelbart WM, Ben-Shaul A (2003) Osmotic shock and the strength of viral capsids. Biophys J 85:70–74
5. Binnig G, Quate CF, Gerber C (1986) Atomic force microscope. Phys Rev Lett 56:930–933
6. Ivanovska IL, de Pablo PJ, Ibarra B, Sgalari G, MacKintosh FC, Carrascosa JL, Schmidt CF, Wuite GJL (2004) Bacteriophage capsids: though nanoshells with complex elastic properties. Proc Natl Acad Sci USA 101:7600–7605
7. Gittes F, Schmidt CF (1998) Thermal noise limitations on micromechanical experiments. Eur Biophys J Biophys Lett 27:75–81
8. Rief M, Oesterhelt F, Heymann B, Gaub HE (1997) Single molecule force spectroscopy on polysaccharides by atomic force microscopy. Science 275:1295–1297
9. Fisher TE, Carrión-Vázquez M, Oberhauser AF, Li H, Marszalek PE, Fernández JM (2000) Single molecule force spectroscopy of modular proteins in the nervous system. Neuron 27:435–446
10. Svoboda K, Schmidt CF, Schnapp BJ, Block SM (1993) Direct observation of kinesin stepping by optical trapping interferometry. Nature 365:721–727

11. Alcaraz J, Buscemi L, Grabulosa M, Trepat X, Fabry B, Farre R, Navajas D (2003) Microrheology of human lung epithelial cells measured by atomic force microscopy. Biophys J 84:2071–2079
12. de Pablo PJ, Schaap IAT, MacKintosh FC, Schmidt CF (2003) Deformation and collapse of microtubules on the nanometer scale. Phys Rev Lett 91:98101
13. Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C (2001) The bacteriophage ϕ29 portal motor can package DNA against a large internal force. Nature 413:748–752
14. Ortega-Esteban A, Horcas I, Hernando-Pérez M, Ares P, Pérez-Berná AJ, San Martín C, Carrascosa JL, de Pablo PJ, Gómez-Herrero J (2012) Minimizing tip-sample forces in jumping mode atomic force microscopy in liquid. Ultramicroscopy 114:56–61
15. Horcas I, Fernandez R, Gomez-Rodriguez JM, Colchero J, Gomez-Herrero J, Baro AM (2007) WSXMa software for scanning probe microscopy and a tool for nanotechnology. Rev Sci Instrum 78:013705
16. Sader JE, Chon JWM, Mulvaney P (1999) Calibration of rectangular atomic force microscope cantilevers. Rev Sci Instrum 70:3967–3969
17. de Pablo PJ, Gomez-Navarro C, Martinez MT, Benito AM, Maser WK, Colchero J, Gomez-Herrero J, Baro AM (2002) Performing current versus voltage measurements of single-walled carbon nanotubes using scanning force microscopy. Appl Phys Lett 80:1462–1464
18. Landau LD, Lifshizt E (1986) Theory of elasticity. Pergamon, London
19. Basak S, Raman A (2007) Dynamics of tapping mode atomic force microscopy in liquids: theory and experiments. Appl Phys Lett 91:064107, 064107-3
20. Israelachvili J (2002) Intermolecular and surface forces. Academic Press, London
21. Klug WS, Bruinsma RF, Michel JP, Knobler CM, Ivanovska IL, Schmidt CF, Wuite GJL (2006) Failure of viral shells. Phys Rev Lett 97:228101
22. Vliegenthart GA, Gompper G (2006) Mechanical deformation of spherical viruses with icosahedral symmetry. Biophys J 91:834–841
23. Zink M, Grubmuller H (2009) Mechanical properties of the icosahedral shell of southern bean mosaic virusa molecular dynamics study. Biophys J 96:1350–1363
24. Ivanovska IL, Miranda R, Carrascosa JL, Wuite GJL, Schmidt CF (2011) Discrete fracture patterns of virus shells reveal mechanical building blocks. Proc Natl Acad Sci USA 108:12611–12616
25. Sokolnikoff IS (1983) Mathematical theory of elasticity. Kriger Pub Co, Florida
26. Greaves GN, Greer AL, Lakes RS, Rouxel T (2011) Poisson's ratio and modern materials. Nat Mater 10:823–837
27. Helfer E, Harlepp S, Bourdieu L, Robert J, MacKintosh FC, Chatenay D (2001) Buckling of actin-coated membranes under application of a local force. Phys Rev Lett 87:088103
28. Niordson FI (1985) Shell theory. North Holland, Amsterdam
29. Hutton DV (2004) Fundamentals on finite element analysis. McGraw Hill, New York
30. Carrasco C, Carreira A, Schaap IAT, Serena PA, Gomez-Herrero J, Mateu MG, de Pablo PJ (2006) DNA-mediated anisotropic mechanical reinforcement of a virus. Proc Natl Acad Sci USA 103:13706–13711
31. Carrasco C, Luque A, Hernando-Pérez M, Miranda R, Carrascosa JL, Serena PA, de Ridder M, Raman A, Gómez-Herrero J, Schaap IAT, Reguera D, de Pablo PJ (2011) Built-in mechanical stress in viral shells. Biophys J 100:1100–1108
32. Schijve J (2009) Fatigue of structures and materials. Kluwer Academic Publishers, Dordrecht
33. Ortega-Esteban A, Pérez-Berná AJ, Menéndez-Conejero R, Flint SJ, San Martín C, de Pablo PJ (2013) Monitoring dynamics of human adenovirus disassembly induced by mechanical fatigue. Sci Rep 3:1434
34. Hernando-Pérez M, Miranda R, Aznar M, Carrascosa JL, Schaap IAT, Reguera D, de Pablo PJ (2012) Direct measurement of phage ϕ29 stiffness provides evidence of internal pressure. Small 8:2366–2370
35. Tao YZ, Olson NH, Xu W, Anderson DL, Rossmann MG, Baker TS (1998) Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell 95:431–437

36. Roos WH, Radtke K, Kniesmeijer E, Geertsema H, Sodeik B, Wuite GJL (2009) Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids. Proc Natl Acad Sci USA 106:9673–9678
37. Gonzalez-Huici V, Salas M, Hermoso JM (2004) The push-pull mechanism of bacteriophage ϕ29 DNA injection. Mol Microbiol 52:529–540
38. Podgornik R, Leforestier A, Siber A, Livolant F (2011) Protein-DNA interactions determine the shapes of DNA toroids condensed in virus capsids. Biophys J 100:2209–2216
39. Purohit PK, Kondev J, Phillips R (2003) Mechanics of DNA packaging in viruses. Proc Natl Acad Sci USA 100:3173–3178
40. Zandi R, van der Schoot P, Reguera D, Kegel W, Reiss H (2006) Classical nucleation theory of virus capsids. Biophys J 90:1939–1948
41. Choi KH, Morais MC, Anderson DL, Rossman MG (2006) Determinants of bacteriophage ϕ29 head morphology. Structure 14:1723–1727
42. Dokland T (1999) Scaffolding proteins and their role in viral assembly. Cell Mol Life Sci 56:580–603
43. Zlotnick A (2003) Are weak protein-protein interactions the general rule in capsid assembly? Virology 315:269–274
44. Mateu MG (2012) Mechanical properties of viruses analyzed by atomic force microscopy: a virological perspective. Virus Res 168:1–22
45. Carrillo-Tripp M, Sheperd CM, Borelli IA, Sangita V, Lander G, Natarajan P, Johnson JE, Brooks CL III, Reddy VS (2009) VIPERdb2an enhanced and web API enabled relational database for structural virology. Nucleic Acid Res 37:D436–D442
46. Carrasco C, Castellanos M, de Pablo PJ, Mateu MG (2008) Manipulation of the mechanical properties of a virus by protein engineering. Proc Natl Acad Sci USA 105:4150–4155
47. Roos WH, Gertsman I, May ER, Brooks CL III, Johnson JE, Wuite GJL (2012) Mechanics of bacteriophage maturation. Proc Natl Acad Sci USA 109:2342–2347
48. Castellanos M, Pérez R, Carrasco C, Hernando-Pérez M, Gómez-Herrero J, de Pablo PJ, Mateu MG (2012) Mechanical elasticity as a physical signature of conformational dynamics in a virus particle. Proc Natl Acad Sci USA 109:12028–12033
49. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJL (2007) Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. Cell Mol Life Sci 64:1484–1497
50. Wilts BD, Schaap IAT, Young M, Douglas T, Knobler CM, Schmidt CF (2010) Swelling and softening of the CCMV plant virus capsid in response to pH shifts. Biophys J 98:656a
51. Roos WH, Wuite GJL (2009) Nanoindentation studies reveal material properties of viruses. Adv Mater 21:1187–1192
52. Roos WH, Bruinsma R, Wuite GJL (2010) Physical Virology. Nature Phys 6:733–743
53. Zandi R, Reguera D (2005) Mechanical properties of viral capsids. Phys Rev E 72:021917
54. Baclayon M, Shoemaker GK, Uetrecht C, Crawford SE, Estes MK, Prasad BVV, Heck AJR, Wuite GJL, Roos WH (2011) Prestress strengthens the shell of Norwalk virus nanoparticles. Nano Lett 11:4865–4869
55. Michel JP, Ivanovska IL, Gibbons MM, Klug WS, Knobler CM, Wuite GJL, Schmidt CF (2006) Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. Proc Natl Acad Sci USA 103:6184–6189
56. Cuellar JL, Meinhoevel F, Hoehne M, Donath E (2010) Size and mechanical stability of norovirus capsids depend on pH: a nanoindentation study. J Gen Virol 91:2449–2456
57. Castellanos M, Pérez R, Carrillo PJP, de Pablo PJ, Mateu MG (2012) Mechanical disassembly of single virus particles reveals kinetic intermediates predicted by theory. Biophys J 102:2615–2624
58. Reddy VS, Johnson JE (2005) Structure-derived insights into virus assembly. Adv Virus Res 64:45–68
59. Singh S, Zlotnick A (2003) Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. J Biol Chem 278:18249–18255
60. Rapaport DC (2008) Role of reversibility in viral capsid growth: a paradigm for self-assembly. Phys Rev Lett 101:1861012008

61. Ivanovska I, Wuite G, Jonsson B, Evilevitch A (2007) Internal DNA pressure modifies stability of WT phage. Proc Natl Acad Sci USA 104:9603–9608
62. Evilevitch A, Roos WH, Ivanovska IL, Jeembaeva M, Jönsson B, Wuite GJL (2011) Effects of salt on internal DNA pressure and mechanical properties of phage capsids. J Mol Biol 405:18–23
63. Eghiaian F, Schaap IA, Skehel JJ, Veigel C, des Georges A (2009) The influenza virus mechanical properties are dominated by its lipid envelope. Biophys J 96:15a
64. Li S, Eghiaian F, Sieben C, Herrmann A, Schaap IA (2011) Bending and puncturing the influenza virus envelope. Biophys J 100:637–645
65. Kol N, Gladnikoff M, Barlam D, Shneck RZ, Rein A, Rousso I (2006) Mechanical properties of murine leukemia virus particles: effect of maturation. Biophys J 91:767–774
66. Kol N, Shi Y, Tsvitov M, Barlam D, Shneck RZ, Kay MS, Rousso I (2007) A stiffness switch in human immunodeficiency virus. Biophys J 92:1777–1783
67. Gelbart WM, Knobler CM (2009) Pressurized viruses. Science 323:1682–1683
68. Reguera J, Grueso E, Carreira A, Sánchez-Martínez C, Almendral JM, Mateu MG (2005) Functional relevance of amino acid residues involved in interactions with ordered nucleic acid in a spherical virus. J Biol Chem 280:17969–17977
69. Reguera J, Carreira A, Riolobos L, Almendral JM, Mateu MG (2004) Role of interfacial amino acid residues in assembly, stability, and conformation of a spherical virus capsid. Proc Natl Acad Sci USA 101:2724–2729
70. Mateu MG (2011) Virus engineering: functionalization and stabilization. Prot Eng Des Sel 24:53–63
71. Johnson JE (2003) Virus particle dynamics. Adv Protein Chem 64:197–218
72. Lander GC, Evilevitch A, Jeembaeva M, Potter CS, Carragher B, Johnson JE (2008) Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location and mechanism of attachment determined by cryo-EM. Structure 16:1339–1406
73. Falvo MR, Washburn S, Superfine R, Finch M, Brooks FP Jr, Chi V, Taylor RM II (1997) Manipulation of individual viruses: friction and mechanical properties. Biophys J 72:1396–1403

Further Reading

Stockley PG, Twarock R (eds) (2010) Emerging topics in physical virology. Imperial College Press, London

Roos WH (2011) How to perform a nanoindentation experiment on a virus. In: Peterman EJG, Wuite GJL (eds) Single molecule analysismethods and protocols, vol 783, Methods Mol Biol. Humana Press, Totowa, pp 251–264

Kurland NE, Drira Z, Yadavalli VK (2012) Measurement of nanomechanical properties of biomolecules using atomic force microscopy. Micron 43:116–128

Also especially recommended for further reading are references [6, 44, 51, 52, 67] *listed above*.

Chapter 19
Theoretical Studies on Assembly, Physical Stability and Dynamics of Viruses

Antoni Luque and David Reguera

Abstract All matter has to obey the general laws of physics and living matter is not an exception. Viruses have not only learnt how to cope with them, but have managed to use them for their own survival. In this chapter we will review some of the exciting physics behind viruses and discuss simple physical models that can shed some light on different aspects of the viral life cycle and viral properties. In particular, we will focus on how the structure and shape of the capsid, its assembly and stability, and the entry and exit of viral particles and their genomes can be understood using fundamental physics theories.

Keywords Biophysics • Virus • Bacteriophage • Capsid • DNA • RNA • Budding • Statistical Mechanics • Continuum elasticity theory • Nucleation • Self-assembly • Wrapping • Pressure • Electrostatics

Abbreviations

AFM	Atomic force microscopy
BMV	Brome mosaic virus
CBB	Capsid building block
CCMV	Chlorotic cowpea mottle virus

A. Luque
Department of Fundamental Physics, Universitat de Barcelona,
c/Martí i Franquès 1, 08028 Barcelona, Spain

Department of Chemistry, New York University, 31 Washington Place,
New York, NY 10003, USA

D. Reguera (✉)
Department of Fundamental Physics, Universitat de Barcelona,
c/Martí i Franquès 1, 08028 Barcelona, Spain
e-mail: dreguera@ub.edu

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*,
Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_19,

CK	Caspar and Klug
CNT	Classical nucleation theory
cryo-EM	Cryo-electron microscopy
DH	Debye-Hückel
dsDNA	Double-stranded DNA
FEA	Finite element analysis
FvK	Föppl-von Kármán number
HBV	Hepatitis B virus
HIV	Human immunodeficiency virus
HK97	Bacteriophage Hong Kong 97
HPV	Human papillomavirus
NMA	Normal mode analysis
ssRNA	Single-stranded RNA
STMV	Satellite tobacco mosaic virus
TMV	Tobacco mosaic virus
TST	Thin shell theory

19.1 Introduction

Viruses are an endless source of fascination for biophysicists. Contrary to most biological organisms, viral particles are made of a minimal number of relatively simple components that are not capable of any metabolic activity, except when their genome takes over the metabolism of the infected host to achieve the replication of new particles (see Chap. 1). Despite the lack of sophisticated biological machinery, viruses have found the way to efficiently infect the host, assemble, and egress the cell following, in many cases, a coordinated sequence of passive and spontaneous processes. This strongly suggests that, during their life cycle, viruses must rely on general physical and chemical mechanisms to succeed in their different tasks and to achieve the required resistance against possible extreme environmental conditions.

The assembly of viruses (Chaps. 10, 11, 12 and 13) provides a good example to illustrate the importance of physical mechanisms in the virus life cycle. In physiological conditions, viral capsids assemble directly from their building blocks, which involve the participation of the capsid proteins, the genetic material, and in some cases, auxiliary proteins. Indeed, it has been shown that many viruses can be reconstituted in a fully infective form by mixing these constitutive elements *in vitro* in the right conditions, as in the paradigmatic case of tobacco mosaic virus (TMV). This means that the interactions among these elements spontaneously drive the formation of viruses in a process that can be described by basic thermodynamic and kinetic principles. Moreover, the physical properties of the viral genome, which differ depending on the type of nucleic acid, have a crucial influence on the viral strategy of assembly and infection. In particular, single-stranded RNA

(ssRNA), which is a relatively flexible macromolecule, usually co-assembles with the viral capsid proteins to form infective viruses. Instead, double-stranded DNA (dsDNA) is a more rigid polymer that many viruses pack at high densities inside a spontaneously preformed shell, the procapsid, using the most powerful molecular motors known in nature. In this way, the elastic energy and high electrostatic self-repulsion of the packed dsDNA phosphate backbones generate an internal pressure that is crucial to initiate the subsequent infection of a new host (see Chaps. 9, 12 and 17).

The wealth of biological details described in previous chapters is essential to confer specificity to viruses and make possible their success infecting a particular host. However, even viruses that infect hosts from different species or kingdoms also share some common physical mechanisms, whose study would pave the way to understand key points in the viral life cycle and to a myriad of potentially useful applications. It is worth noting that viruses have been subjected to natural selection, some of them for billions of years, perhaps since the origin of life. Thus, they constitute a pool of optimal solutions at the nanoscale of incalculable value, and their efficiency in performing all kinds of functions make them a perfect model to guide the design of a new generation of nanostructures. In fact, the impressive properties of viral capsids have already grounded the development of promising biomedical and nanotechnological applications, as described in Chaps. 20, 21 and 22.

In this chapter, we will review the physics behind some of the major different steps of the virus infection cycle, and we will introduce the main theoretical tools developed in the emerging field of physical virology. In particular, we will discuss some of the underlying physical principles and likely mechanisms involved in virus architecture, viral particle assembly and mechanical stability, nucleic acid packaging, and virus exit from the infected cell and entry into another host cell; and we will see how simple theoretical models are able to give a very useful insight into many aspects of viruses, despite not incorporating most of the structural and biological elements introduced in other chapters of this book.

More specifically, first, we will see how the exquisite regular shape and architecture of viral particles is a natural consequence of simple geometrical principles and free energy minimization. Then, we will discuss how the favorable interactions among capsid proteins lead to the spontaneous formation of capsids in a self-assembly process that can be described using the same models that explain the phase transformations of well-known non-biological systems. The presence and the physical properties of the genome play a central role in this process, hence ssRNA and dsDNA viruses will be discussed separately. Once assembled, the physical stability of viruses is crucial to ensure their infectivity, and this stability is related to the mechanical properties of the capsid. Classical elasticity theory will be the starting framework to study capsid mechanical stability and, complemented with molecular modeling, will provide essential information to understand the physical resistance of viral shells. Finally, in the last section of this chapter, we will describe some basic physical mechanisms behind genome delivery and virus egress strategies during the infection cycle.

19.2 Architecture of Viral Shells

An astonishing aspect of viruses is the well-defined shape and symmetry of their capsids. As discussed in Chap. 2, viral shells essentially adopt four different architectures: quasi-spherical, rod-like, bacilliform, and, more rarely, conical. Thus, completely unrelated viruses can form similar structures, which is truly remarkable, specially taking into account that they generally infect different hosts and are made out of capsid proteins whose amino-acid sequences, sizes, or conformations can be very diverse. This suggests that, underlying the virus-specific structural details, there must be some common physical and geometrical principles dictating the shape and architecture of capsids.

The aim of this section is to provide a brief overview of the main theoretical ideas and models that explain the architecture of viral capsids. They have been grouped in three categories. First, the geometrical concepts that justify the overall shape and symmetry of viruses are discussed in Sect. 19.2.1. These geometrical principles provide a catalogue of the possible architectures adopted by viral capsids using very precise rules of construction. However, these ideas do not suffice to explain why a specific virus prefers a particular architecture. The selection of concrete structures obeys the physical principle of free energy minimization, and several models of increasing level of complexity have been proposed to understand the details of this process. The simplest models describe viral shells as continuum elastic materials, and their description in terms of thin shell theory (TST) is introduced in Sect. 19.2.2. Nevertheless, the fundamental understanding of architectural details of capsids requires the use of discrete models, and in Sect. 19.2.3 we discuss simulations of coarse-grained and discrete biomolecular models that allow exploring more complex and realistic scenarios. The combination of continuum and simulation approaches is fundamental to establish a solid theoretical framework that can rationalize our understanding of viral structures, helping us to guide further experiments and applications.

19.2.1 Geometrical Theories

One can justify the construction of the four main morphologies of viral capsids using basic geometrical ideas. As mentioned in Chap. 2, the small size of viral capsids, on the order of tens to hundreds of nanometers, restricts the amount of information that can be coded in the viral genome contained inside. Therefore, for the sake of genetic economy, capsids are typically built from multiple copies of a single or a few different small proteins [1]. These identical subunits interact with each other building a regular structure, in a process that shares some similarities with the crystallization of molecules that form regular lattices [1, 2]. However, in the case of viruses, the capsid proteins construct a hollow shell rather than a three dimensional solid. In a planar surface, the organization allowing the highest

packing of identical units, which maximizes the number of interactions, is the hexagonal lattice or, equivalently, the triangular lattice. Now, starting from this hexagonal network it is possible to rationalize the construction of all basic capsid shapes. In particular, the open helical tubes typical of rod-like viruses can be obtained by simply wrapping up the lattice. Alternatively, to obtain a closed shell, one needs to introduce 12 pentagonal defects (see Chap. 2). If those are evenly distributed, one gets an icosahedral shell, characteristic of quasispherical viruses. In a similar way, prolate or bacilliform capsids can be obtained by wrapping the lattice into a helical tube and closing each of its ends with hemispheres containing six defects. Finally, conical viruses are obtained by making a tube with a conical section closed with a different number of defects at the ends.

The geometrical characterization of viral capsids is particularly important because there is only a limited number of ways to introduce the pentameric defects that lead to perfectly closed regular shells. For spherical viruses, the different possibilities give rise to the triangulation T-number used in the structural classification of icosahedral capsids (see Chap. 2). For bacilliform viruses, the symmetry and arrangement of the pentameric defects in the caps determine a discrete set of compatible lengths, radii, and helicities that can be adopted. Finally, the different distribution possibilities of the 12 pentameric defects in both ends of a conical shell, *e.g.*, 7 and 5, or 9 and 3, explain the discretized values of cone angles found in viruses like human immunodeficiency virus (HIV). Indeed, these simple geometrical rules provide the basis for the structural classification of spherical viruses introduced back in the 60s by Caspar and Klug (CK), which was extended for some bacilliform viruses by Moody and has been generalized recently [3] (see Chap. 2).

Graphically, these structures can be considered as the result of covering a sphere, a spherocylinder, or a conical tube using equilateral triangular tiles that are associated to the capsid subunits (see Fig. 19.1a, b). Nevertheless, certain viruses do not fit strictly into this classification. For instance, polyomavirus has a structure similar to a $T = 7$ capsid where the expected hexagonal and pentagonal positions in the icosahedral lattice are occupied by clusters of five proteins (see also Chap. 2). To tackle this puzzling situation, an alternative tiling theory[1] has been developed recently, introducing tiles of other shapes as rhombs or kites rather than the triangular tessellation of the CK theory (see Fig. 19.1c). These ideas have been successful in characterizing the structure of some spherical and tubular all-pentamer viruses [5]. More recently, the same group showed that structural details on the tertiary structure of viral proteins, and even the organization of the genomic material inside the capsid could be reproduced using a sophisticated extension of the icosahedral symmetry, which works as a geometrical blueprint [15].

Therefore, all these geometrical models are very helpful characterizing the structure of viral capsids. Nonetheless, they do not explain *why* viruses adopt those particular architectures. In this sense, several simple mathematical models

[1] Tiling theory is a set of mathematical operations which can be used to cover an arbitrary surface with geometrical shapes (or tiles) with no overlaps and no gaps.

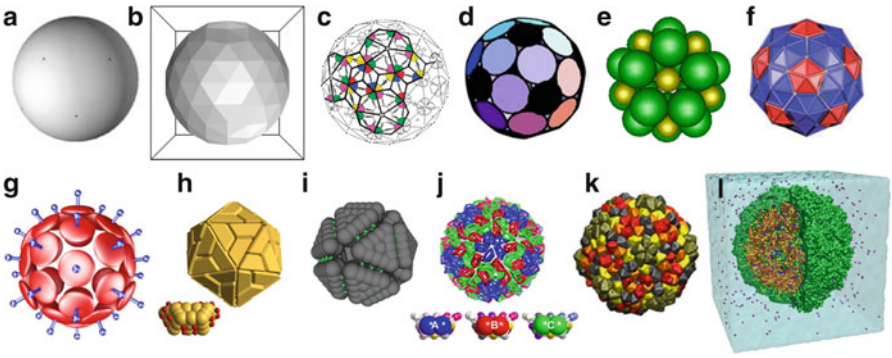


Fig. 19.1 Different levels of modelization of viral capsids. (**a**) Continuum elastic model of a spherical capsid in TST. (**b**) Representation of a shell as a triangulated network [4]. (**c**) Tiling representation of polyomavirus using kite and rhombic tiles [5]. (**d**–**g**) Coarse-grained models where capsomers and proteins are represented as (**d**) disks [6] or (**e**) balls [7] over the surface of a sphere, (**f**) hexagonal and pentagonal pyramids [8], and (**g**) ellipsoidal capsomers with additional attractive and repulsive sites [9]. (**h**–**j**) Coarse-grained models at the level of proteins described as: (**h**) multiple bead trapezoidal objects with many interaction sites (shown below) [10], (**i**) truncated pyramids [11], (**j**) trapezoidal multi-beads of three different types (shown below) [12]. (**k**) Shape-based coarse-grained model of phage ΦX174 [13]. (**l**) Full-atom model, including water and ions, of satellite tobacco mosaic virus (STMV) [14]

have tried to justify the prevalence of these structures in terms of their possible energetic or biological advantages. One of these desirable features is storage efficiency, which could be translated into maximizing the volume-to-surface ratio, *i.e.*, maximizing the storage capacity of the capsid with the minimum amount of proteins. The first ideas in this direction were based on exploring models that maximized the number of contacts, the packing, or the covering of a sphere by simple geometric objects like disks, which represented the capsomers or the structural subunits of the capsid [16]. These models were able to reproduce the symmetries of some all-pentamer viruses, but did not recover the sequence of T-numbers described by CK and adopted by most spherical viruses. Interestingly, the mathematical ideas of packing and covering are somehow connected to a more general principle in physics: the free energy minimization. This idea, already suggested by CK [2], has been explored using different continuum and discrete models as described in the next two subsections.

19.2.2 Continuum Theories

A first approximation to the energetic description of different capsid shapes can be made using the well-known continuum elasticity theory [17]. In this framework, the viral capsid is described as a continuum homogeneous material, and the focus is placed on the elastic energy required to deform the shell in order to obtain the

different capsid shapes. At first sight, this continuum description might seem very crude, since viruses are relatively small and made out of discrete elements (proteins). However, even the smallest virus contain hundreds of thousands of atoms that render this approximation not so unrealistic. Moreover, in most viruses the radius of their protein shell is substantially larger than its thickness, which means that the capsid can be approximated as a two-dimensional surface, and described using a simplified limit of the elastic theory known as thin shell theory (TST). This approach has been already used to describe the structural properties of several biological systems, like membranes and vesicles, and constitutes a very convenient starting point to understand the energetics of viral capsids.

Thin Shell Theory (TST)

In the context of TST, the elastic energy of a viral shell is decomposed into two main contributions [4]. First, one has a *stretching* term that is the energetic cost associated to elongating or compressing a planar piece of capsid material. This cost is commonly characterized by two elastic constants: the two-dimensional (2D) Young's modulus Y, which describes the stiffness of the material and is given by the ratio between the applied stress and the resulting elongation per unit length; and the Poisson ratio ν, which describes the relative contraction in the transverse directions when a material is stretched in the longitudinal direction (see Chap. 18). For many protein-like materials, this is taken to be around $\nu \sim 0.3$ or 0.4, which is similar to rubber. The second energetic contribution is the *bending* term, which accounts for the out-of-plane deformation. This is characterized by the bending constant κ, which describes the resistance to flexion, and a spontaneous curvature C_0, which is the inverse of the preferred radius of curvature of the surface. It is worth noting that C_0 is ultimately defined by the directionality of the interactions between capsid proteins and the steric effect of their surfaces in contact. Additionally, there is a contribution associated with the Gaussian rigidity, which for closed surfaces with fixed topology, like a sphere, a spherocylinder, or a cone, is a constant that is generally not taken into account.

Furthermore, as discussed in Chap. 2, spherical, bacilliform, and conical viruses require the insertion of 12 pentameric defects, or *disclinations*, in the original planar hexagonal network to form a closed shell. These disclinations are made of the same material as the rest of the capsid and generally accumulate an important degree of stress, which grows with the size of the shell and can promote the faceting of the structure, as we will see in Sect. 19.4.2. The contribution of these disclinations is implicitly captured in the TST description, although explicit expressions can be also derived. Interestingly, the total elastic energy of a shell obtained by adding up all these contributions depends only on a single non-dimensional parameter known as the Föppl-von Kármán (FvK) number, $\gamma = YR^2/\kappa$, where R is the radius of the shell. The FvK number is essentially the ratio of the stretching and bending energies and considerably simplifies the study of the capsid properties in the context of TST [4, 18, 19].

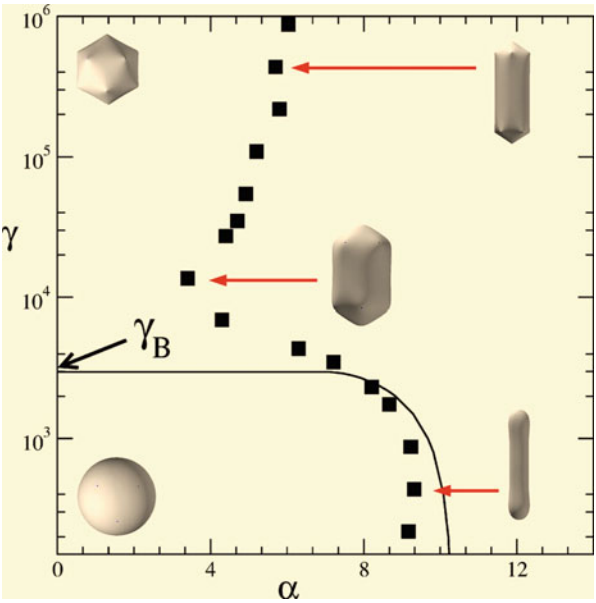


Fig. 19.2 Phase diagram of capsid shapes predicted by continuum theories (TST). Diagram showing the optimal shapes (*i.e.*, those that minimize the elastic energy in the context of TST) as a function of the FvK number, $\gamma = YS/\kappa$, and the dimensionless spontaneous curvature $\alpha = C_0 S^{1/2}$, where S is the area of the capsid. For low values of the FvK number, *i.e.* $\gamma < \gamma_B$, the optimal shape is spherical at low spontaneous curvatures and spherocylindrical, at high α. For $\gamma > \gamma_B$, a buckling transition is predicted and the optimal shapes are increasingly faceted spheres or spherocylinders. The continuous line is the TST analytical prediction, and the squares are the numerical results obtained from minimization of the elastic energy using a discretized triangular network. (Reproduced with permission from Ref. [18])

As has been discussed above, the particular shape and structure of viral capsids is the consequence of the free energy minimization principle. This means that, to obtain the possible architectures of viral capsids within TST, one has to minimize the energy functional containing the stretching and bending energies, including the contribution of the disclinations. This can be challenging in many cases, but, rather than solving the analytical expressions of TST, one can also minimize the elastic energy of a shell in a more convenient way: discretizing the structure into a triangular network and performing numerical simulations using the so-called *mass-and-spring model*, which approximates the interactions by a set of harmonic springs [18].

In this way, it has been possible to investigate the phase diagram of the optimal viral shapes as a function of the spontaneous radius of curvature and the FvK number [18] (see Fig. 19.2). The diagram shows that spherical capsids are more stable than spherocylindrical shapes at low values of the spontaneous curvature for a fixed surface, *i.e.*, number of capsid proteins. However, conical capsids were never found to be energetically better than spherical or spherocylindrical capsids. This suggests that in the context of TST some alternative physical constraints are

necessary to explain the existence of conical architectures, such as those adopted by the mature capsid of HIV-1 (see Chap. 2). In addition, for large FvK, capsids prefer to adopt a relatively faceted shape in order to release the stress accumulated in the disclinations, in a process known as the buckling transition [4], which has been well reported in the maturation of quasi-spherical capsids (see Sect. 19.4.2).

Thus, TST provides a very useful starting point to explain the overall shapes of viral capsids. However, TST does not generally capture the possible distribution of capsomers or capsid subunits in a shell. This means that in the context of continuum elastic theories there is not a simple way to distinguish between two quasispherical capsids with different T-numbers. To overcome this problem it is more convenient to use discrete models, as discussed below.

19.2.3 Coarse-Grained, Discrete Models, and Simulations

The particular structure adopted by a viral capsid is ultimately a consequence of the interactions between its fundamental building blocks. In the framework of TST, the molecular interactions between capsid proteins are effectively replaced by the stretching and bending elastic terms. However, to explore equilibrium and dynamical properties of capsids in a higher degree of detail, one cannot ignore the discrete nature of viral shells.

In this context, the discrete modeling of capsids has been performed at different levels of resolution, from full-atom approaches to simple coarse-grained physical models (see Fig. 19.1). The appropriate choice of the resolution depends on different factors, like the properties to be probed, the existence of experimental data, and the computational limitations. Here we have organized these modeling strategies in two groups. The first one contains *low-resolution* models that are usually grounded in basic physical ideas, aiming to understand the essential ingredients required to explain common aspects of viral structures; the second group includes *high-resolution* models, which generally use specific molecular reconstructions derived from experiments as a starting point, and then introduce force fields that can vary in complexity in order to properly characterize the finer structural details of particular virus capsids.

Low-Resolution Models

Low-resolution models might seem a rough approximation to describe complex molecular structures like capsids, but they constitute an excellent strategy to improve our understanding of viruses. The absence of molecular details allows the simulation of relatively large systems, and the simplicity of the models facilitates the qualitative characterization of the underlying mechanisms that explain many important properties of viral capsids. In general, these models replace the stretching and bending energies, discussed above in the context of continuum theory, by several types of effective interactions between the capsid building blocks

(CBBs). Three are essentially the minimal ingredients of these interactions. First, one needs to include a short-range repulsion force that mimics the steric effects between capsid proteins, which is intended to avoid the overlapping between the molecules. Then, a longer-range attraction is also introduced to drive the self-assembly of the capsid and to give stability to the final structure. The third essential ingredient is the existence of an imposed curvature, or preferential angle of interaction, between the building blocks, which is ultimately responsible for the selection of a particular size and curvature of the shell.

Remarkably, the seminal work in Ref. [6] proved that the icosahedral and other symmetrical structures adopted by spherical viruses are the natural consequence of the free energy minimization of a very generic interaction, which essentially contains the contributions discussed above. In their low-resolution model, the description was made in terms of capsomers, represented as disks or spheres of two different radii (see Fig. 19.1d), corresponding to pentamers and hexamers, which were confined to the surface of a sphere. Then, using a computational technique known as Monte Carlo simulation, it was possible to find the optimal arrangement of capsomers. For the case where hexamers and pentamers coexist in the capsid, the distribution that minimized the total energy corresponded precisely to the T-number architectures of the CK classification (see Fig. 19.3). Additionally, when only one type of morphological unit was allowed, the optimal free energy minima corresponded to the icosahedral and non-icosahedral architectures observed in all-pentamer viruses both *in vivo* and *in vitro*.

More recently, this model was extended to study the optimal structures of bacilliform capsids [7] (see Fig. 19.3c, d) showing that icosahedral symmetry is also predominant on the ending caps of these shells. Interestingly, the optimal structures followed the set of geometrical rules derived in Ref. [3] as extension of CK for elongated viruses. In addition, this study justified the existence of non-icosahedral-capped capsids for all-pentamer viruses (*e.g.*, polyomavirus), and demonstrated that the bacilliform geometrical construction also captures most of the properties of icosahedrally-capped all-pentamer viruses.

The important message of these simple coarse-grained models is that the impressive regular and symmetric architectures of viral capsids emerge naturally from free energy minimization of very generic interactions between their building blocks. In other words, the details of these interactions turn out to be not so crucial in determining the possible catalogue of viral shell architectures, although they can be essential to determine the physiological conditions of capsid formation. In fact, various models that use different types of interaction potentials and more realistic representations lead to similar conclusions [9, 21]. The exhaustive thermodynamic exploration of these models has shown that the set of basic parameters introduced in these simple physical approaches are able to spontaneously reproduce spheroidal shells, tubular and helical shapes, and even head-tail morphologies [9]. In this way, these low-resolution strategies have played a very important role in unveiling the key factors that determine the architecture of viral capsids. Furthermore, they are contributing to the determination of fundamental mechanisms controlling the mechanical properties, and the assembly and disassembly of viral capsids either in the presence or absence of genetic material, as will be discussed in Sects. 19.3 and 19.4.

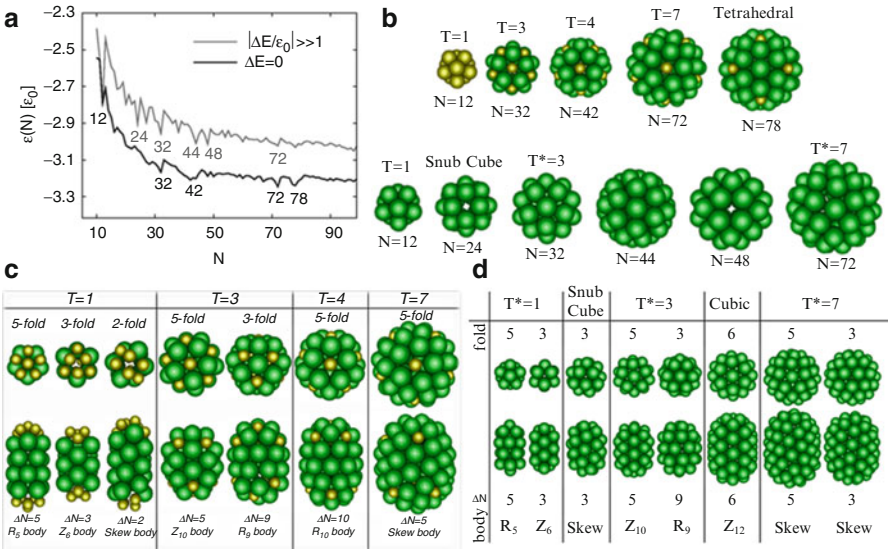


Fig. 19.3 Gallery of optimal structures obtained from coarse-grained simulations. (**a**) Energy landscape of the optimal structures obtained in a coarse-grained model of a quasi-spherical capsid with two types of capsomers (*solid line*), *i.e.*, pentamers and hexamers, or just one type (*grey line*). The figure plots the average energy per capsomer as a function of the total number of capsomers of the shell, and the deepest minima correspond to the optimal structures, which for the two capsomer-types coincide with the CK T-number structures. Notice that the apparently small energy differences *per capsomer* plotted correspond to quite large energy differences *per capsid*. (**b**) Gallery of optimal structures for quasispherical viruses made of two types of capsomers, *i.e.*, pentamers (*gold*) and hexamers (*green*) (*top*), or just of one type (*bottom*). (**c** and **d**) Gallery of optimal structures for spherocylindrical viruses with two (**c**) and just one (**d**) capsomer types. The top images show a zenithal view of the caps, and the lower images show a lateral view of the complete bacilliform virus. In each case, the T-number of the cap, the minimum number of capsomers in the body *ΔN* required to elongate the structure, and the arrangements of capsomers in the body (forming rings R_n or zig-zag patterns Z_n of n capsomers) are also indicated. (Figures adapted from Refs. [7, 20])

High-Resolution Models

High-resolution models are the most appropriate computational strategy to study in detail the properties of specific viruses. Contrary to the models discussed above, these are based on structures reconstructed from X-ray crystallography (see Chap. 4) and/or cryo-electron microscopy (cryo-EM; see Chap. 3) studies. The interactions between the atoms are modeled through effective force fields, and techniques like molecular dynamics or Monte Carlo are applied to sample the structural dynamics of the viruses.

Full-atom simulations are gaining popularity in the field of structural and chemical biology, and are of particular interest for biomedical applications. The constant improvement of computational power and sampling techniques has made this approach more feasible in the last years, though there is still ongoing research

to develop more accurate force fields and to reach longer time scales in simulations. The first full-atom simulation with explicit solvent of a complete virus was developed for the relatively small satellite tobacco mosaic virus (STMV), involving a million of atoms (135,960 protein atoms, 30,330 nucleic acid atoms, plus water and ions) and achieving timescales of nanoseconds. In that case it was shown that, at physiological conditions, the virus becomes unstable when the viral RNA is removed [14]. To study larger systems or processes that involve longer time scales it is required to use multiscale, or mesoscale models, and some degree of coarse-graining [22].

These simulations are starting to provide very useful insights into the detailed structural mechanisms of different viruses. For instance, in the case of HIV, it has been possible to study the interactions between the main capsid protein and the lipidic membrane in the immature virus [23], the role of the C-terminal binding interface between the capsid proteins controlling the curvature and shape of the capsid [23], and the phase diagram for the assembly of the capsid proteins [24]. Another well-studied case has been the plant virus cowpea chlorotic mottle virus (CCMV), where all-atom simulations have characterized its swelling process that is triggered by changes in the pH [25]. In particular, the combination of elastic network with coarse-grained models has allowed simulating the structure of this virus in the scale of microseconds. Besides, general multiscale models have been applied to simulate the *Nudaurelia capensis* ω virus and human papillomavirus (HPV). Alternatively, coarse-grained simulations of the bacteriophage MS2 indicates that the organization of the genetic material within the capsids of RNA viruses depends on the non-specific self-repulsion of the RNA in the inner layers, but the organization of the outer layers strongly depends on the interactions with the inner part of the capsid. Finally, simulations of the structure of Pariacoto virus suggest assembly mechanisms of RNA viruses in agreement with alternative physical models of assembly [26] (see Sect. 19.3).

19.3 Assembly of Viruses

The assembly of viral particles is a crucial step in the life cycle of viruses. Depending on the virus, the capsid can successfully assemble from the capsid proteins alone, or in combination with scaffolding proteins, or the genetic material, following specific assembly pathways reviewed in Chaps. 10, 11, 12 and 13. However, the assembly of viruses also shows some general trends that seem susceptible to be modeled using basic physical ideas. As discussed in Sect. 19.2, a very important fact is the spontaneity of the self-assembly process of viruses with single-stranded genomes, which is driven by free energy minimization, meaning that no external energy in form of ATP is required. This makes the proper assembly of some viral shells possible even in the absence of genetic material or in the presence of other types of cargoes. In fact, many viruses can be reconstituted *in vitro* from their basic components, yielding particles that are fully infective and indistinguishable from

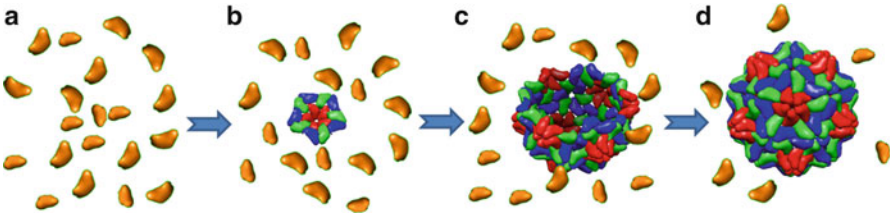


Fig. 19.4 Sketch of the different steps in capsid assembly. (**a**) The process is initiated by the presence of a sufficiently high concentration of free capsid building blocks (in *gold*). (**b**) These building blocks aggregate and disaggregate by thermal fluctuations until a sufficiently large cluster of subunits (in *color*) forms, which acts as the *nucleus* or embryo of the capsid; (**c**) the embryo grows by the addition of building blocks and forms a partial capsid or intermediate of *n* building blocks; (**d**) finally a complete capsid is formed

native virions (see Chaps. 10 and 21). The best-known example is TMV, which was the first virus to be fully reconstituted *in vitro* by mixing the capsid proteins and the viral RNA. In addition, by properly tuning the assembly conditions it has been possible to produce empty capsids of viruses that *in vivo* require the genetic material to assemble, as has been shown for plant viruses like CCMV and brome mosaic virus (BMV), or animal ones like human polyomavirus and hepatitis B virus (HBV). The modification of assembly conditions, like capsid protein concentration, pH, or ionic strength, can also lead viral capsid proteins to assemble into capsid structures different from the native one [27], and to encapsulate foreign genetic materials and even non-organic cargoes [28, 29]. Obviously, this has opened the possibility to engineer viral capsid proteins for different applications in nanoscience, *e.g.*, encapsulation of materials or templating of nanostructures (see Chap. 22).

Diverse viruses can follow different assembly pathways (see Chaps. 1, 10, 11, 12 and 13). In particular, their building blocks (CBBs) can be made of different number of capsid protein subunits and are typically defined as being the most stable oligomer in solution. For instance, the CBBs are single proteins in the case of *Penicillium chrysogenum* virus, dimers for CCMV, pentamers in HPV, or pentamers and hexamers in solution for bacteriophage Hong Kong 97 (HK97). Starting from these building blocks in solution, the assembly process for empty capsids proceeds first by the formation of some aggregates of CBBs, known as *nuclei*, which progressively grow incorporating more CBBs, forming what will be called *partial capsids* or *intermediates*, until a complete shell is assembled (see Fig. 19.4). It is worth noticing that understanding the steps during viral assembly is a key question to identify, for instance, possible targets for antivirals (see Chap. 20).

From a theoretical point of view, most modeling efforts have focused on the simplest case, *i.e.*, the *in vitro* self-assembly of *empty* viral capsids from capsid proteins that do not require any sort of auxiliary proteins. The abundance of *in vitro* experiments has facilitated the investigation of this scenario, and it is the first topic to be described in this section. Subsequently, we will discuss separately the peculiarities of the assembly of ssRNA and dsDNA viruses, since the mechanism followed in the assembly of nucleocapsids depends crucially on the type of nucleic acid of the viral genome.

19.3.1 Empty Capsids

The fact that many viruses can assemble in the absence of genetic material indicates that the interactions between viral capsid proteins generally contain all the information to form a capsid. The range and specificities of these interactions can vary among viruses, but they are usually dominated by two non-specific contributions: hydrophobic and electrostatic [30].

Hydrophobic and Electrostatic Interactions

Many viruses, whose structure has been solved close to atomic resolution, have interfaces between subunits with a strong component of hydrophobic amino acid side chains. In addition, structural and biochemical studies with some model viral particles have shown that the strength of viral capsid protein-protein interactions tends to increase when the temperature is raised, indicating that hydrophobic interactions are the main driving force in the assembly of those particles. The hydrophobic effect is ubiquitous in biological systems, although it is still a phenomenon under intense investigation. However, some effective characterizations have been already tested in non-biological soft matter systems, which conveniently serve as a starting point for the study of viruses.

In particular, the effective hydrophobic interaction potential between two capsid subunits can be approximated by that of two identical apolar surfaces, given by $V_H = -2\gamma A \exp(-D/\gamma)$. In our case γ denotes an effective surface tension between the hydrophobic protein surface and water, A the contact area between capsid proteins, D the separation of the hydrophobic surfaces, and λ a decay length, which is typically on the order of nanometers. The distance between interacting subunits in a capsid is on the atomic scale, *i.e.*, $D << \lambda$, and so the exponential term in the final capsid will be close to unity. Thus, for a complete capsid the total hydrophobic contribution to the free energy of formation will be $G_H = -A_H\gamma$, where A_H is the total hydrophobic area buried in the formation of the shell.

In addition, electrostatics plays an important role in many viral capsids. Capsid proteins are in many cases rich in basic residues leading to an excess of positive charge in the interior of viral shells. This allows the subunits to interact with the negatively charged phosphate backbone of the viral nucleic acids, promoting the assembly and stability of the capsid. In the absence of the genetic material this overcharge introduces an effective repulsion between the capsid proteins. As a consequence, the assembly of empty capsids may require higher salt concentrations (to screen the electrostatic repulsions) and/or lower pHs (to reduce the charge of protein residues), compared with the co-assembly with the genetic material. In addition, in several capsids, electrostatic repulsions between either positively or negatively charged residues that may occur under certain conditions *in vivo* are

used to regulate capsid conformational stability, disassembly or genome uncoating (*e.g.*, in TMV, CCMV and foot-and-mouth disease virus).

In a first approximation, the electrostatic contribution in a viral capsid can be described using Debye-Hückel (DH) theory [31]. This takes into account that the ions of the salt in the medium will screen the electrostatic interaction between the charged residues of a capsid. For two unit charges e separated by a distance r this description leads to the potential $V_c = k_B T(\lambda_B/r)\ \exp\ (-\kappa r)$, where k_B is the Boltzmann constant, T the absolute temperature, $\lambda_B = e^2/4\pi\varepsilon_0\varepsilon_r k_B T$ the Bjerrum length, and $\kappa^{-1} = \left(4\pi\lambda_{\mathrm{B}} \sum_i z_i^2 c_0^i\right)^{-1/2}$ the Debye screening length. Here $\varepsilon_0\varepsilon_r$ is the permittivity of the medium (water in this case), c_0^i is the number density of charged species i in the electrolyte (in units of m^{-3}), and z_i their corresponding valence. In the crudest approximation, a quasi-spherical capsid can be considered as a uniformly charged thin sphere of radius R and surface charge density σ, surrounded by an (ideal) ionic solution (1:1 electrolyte) of bulk concentration c_0. The electrostatic contribution to the free energy of the capsid is then obtained by integrating the DH potential over the surface of the sphere, yielding $G_{elec} = \frac{\pi\sigma^2 R^2}{\varepsilon_0\varepsilon_r\kappa}$, provided that the Debye length is short enough, *i.e.*, $\kappa R >> 1$, which is usually satisfied even at low salt concentrations, *e.g.*, on the order of mM (for a more detailed derivation of both terms, the reader is referred to Refs. [30, 32]). This simple approximation can be refined by incorporating the effects of a finite capsid thickness, a nonuniform distribution of charges in the capsid, lower ionic strengths, or even by solving the more rigorous description of this electrostatic problem given by the Poisson-Boltzmann equation. Interestingly, all these approaches lead to qualitatively similar results [32].

In this context, the total free energy of a complete spherical capsid is obtained by adding the hydrophobic and electrostatic contributions, yielding $G \approx -A_H\gamma + \frac{\pi\sigma^2 R^2}{\varepsilon_0\varepsilon_r\kappa}$. Rough estimates of the order of magnitude of both contributions can be made using HBV as an example. For HBV, the total hydrophobic buried surface is $A_H \sim 10^3\ \mathrm{nm}^2$, and the energy per unit area of exposed protein (accounting also for van der Waals interactions) is roughly $\gamma \sim 6\ \mathrm{mJ/m}^2$, yielding an estimate of $\sim 10^4\ k_B T$ as a typical attractive contribution. The order of magnitude of its surface charge is $\sigma \sim 0.7\ \mathrm{e/nm}^2$, and its radius is $R \sim 30$ nm, giving also an electrostatic contribution of $\sim 10^4\ k_B T$ at physiological conditions ($c_0 = 150$ mM) where $\kappa^{-1} \sim 1$ nm. Both contributions have thus similar strengths, illustrating the delicate balance between the two required for a successful assembly and its sensitivity to the environmental conditions.

The main parameter controlling the strength of the hydrophobic interaction is the temperature, whose increase favors a stronger attraction. In addition, the electrostatic contribution can be tuned by two accessible physicochemical conditions: the pH and the salt concentration of the solution. The pH controls the fraction of charged residues in the capsid protein, and the salt concentration varies the ionic strength, which screens the electrostatic repulsion between capsid proteins. These three factors have major roles in controlling the feasibility of capsid assembly both *in vivo* and *in vitro*.

Capsid Assembly as a Thermodynamic Process

The fourth important factor in the self-assembly of capsids is the concentration of subunits, and its impact on the assembly of empty capsids has been probed *in vitro* in the last years. In those assembly experiments, one starts from a given concentration of capsid proteins in solution and monitors the formation of capsids by scattering techniques, electron microscopy, or size-exclusion chromatography. These studies have revealed a steep dependence of the assembly yield on protein concentration as shown in Fig. 19.5a for the paradigmatic case of HBV. For low protein concentrations capsids are not formed; then the production starts above a certain threshold concentration, and grows steeply with concentration, until eventually reaching saturation. Another distinguishing feature observed experimentally is the lack of intermediates, *i.e.*, most subunits are found either free in solution or in fully formed capsids.

This behavior can be described theoretically by using standard ideas from equilibrium thermodynamics. The process of assembly can be thought as a phase transition between two states: the capsid building blocks in solution and in a fully formed capsid. The equilibrium state is reached when the chemical potentials of both states are equal, leading to the well-known law of mass action

$$K_{capsid}^{eq} = \frac{c_{capsid} c_s^{q-1}}{c_1^q} = e^{-\frac{G}{kT}}, \tag{19.1}$$

where q is total number of subunits in a complete capsid, c_{capsid} and c_1 are the equilibrium concentrations of capsids and subunits, respectively, c_s is the molarity of pure water taken as a reference concentration, and G is related to the free energy of capsid formation described above. The law of mass action can be conveniently rewritten in terms of the capsid mass fraction $f \equiv qc_q/c$ as

$$1 - f = f^{1/q} \left(\frac{c^*}{c} \right)^{1-1/q} \approx \begin{cases} 1, & c < c^* \\ c^*/c, & c \geq c^* \end{cases}, \tag{19.2}$$

where c is the total concentration of proteins in the solution, and $c^* = c_s e^{\frac{\Delta g}{kT}}$ is a threshold concentration related to the binding energy per protein $\Delta g \equiv G/q$. In terms of this threshold concentration, the final yield of capsid production can be described by a universal curve given by Eq. (19.2). Interestingly, since for viruses one typically has $q >> 1$, this universal curve can be simplified to $f = 0$ for $c < c^*$ and $f = 1 - c^*/c$ for $c \geq c^*$. This simple model successfully described the temperature and salt concentration dependence of the *in vitro* assembly of HBV [30], as shown in Fig. 19.5b.

Kinetic Aspects of Viral Assembly

In the past years, different studies using light scattering and turbidimetry techniques have been able to monitor also the kinetics of the assembly process for empty

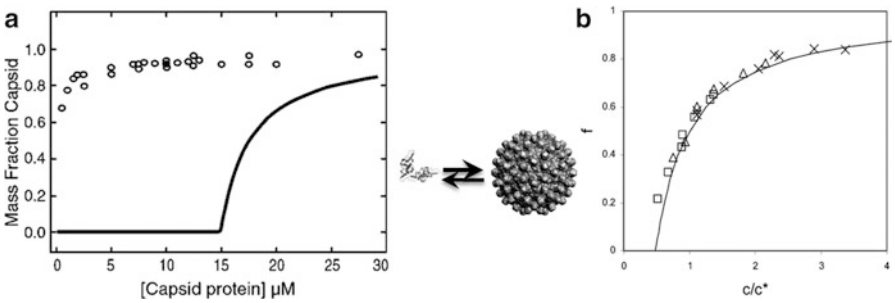


Fig. 19.5 *In vitro* assembly of HBV capsids. (**a**) Mass fraction of proteins in capsids as a function of the total capsid protein concentration in assembly (*solid line*) and disassembly (*open circles*) experiments. (**b**) Fraction of subunits in complete capsids, *f*, *versus* the total concentration of capsid protein scaled by the critical micelle concentration $c_{cmc} = 2c^*$. The symbols are data from Ref. [33] obtained from experiments of HBV assembly at 25 °C and different buffer concentrations of NaCl (crosses $c_{salt} = 0.7$ M; triangles $c_{salt} = 0.5$ M and squares $c_{salt} = 0.3$ M). The line is the universal aggregation curve predicted by Eq. (19.2). (Adapted from Ref. [30, 34])

capsids like HBV or HPV. These experiments suggest again the existence of common features in the kinetics of capsid assembly. It turns out that initially there is a lag time period where no capsids are formed, and then the production rapidly increases following a sigmoidal curve. Another characteristic is that the disassembly of the capsid takes place at different concentration conditions than the assembly, giving place to a hysteresis phenomenon, which is observed even in viruses that do not go through any apparent post-assembly maturation (see Fig. 19.5a). Indeed, once capsids are formed, they do not dissociate easily even if the concentration of free subunits is drastically reduced. This has a clear biological relevance, since viruses must resist a number of different environments to reach a new host, where no viral proteins are present.

Again, these features match the expected behavior of a standard first-order phase transition. Accordingly, different models have borrowed the physical concepts involved in the kinetics of phase transitions to quantify and characterize the assembly of viral capsids.

The first kinetic model for viral assembly was introduced by Zlotnick and collaborators [35]. There, the assembly of the capsid is conceived as a series of concurrent molecular reactions between individual subunits and intermediates of all possible sizes, which are schematically represented as subunits ⇔ nuclei + subunits ⇔ capsids (see Fig. 19.4). The population of intermediates containing n subunits is given by the set of equations

$$\frac{d[n]}{dt} = k_{elong,n-1}[n-1][subunit] + k_{dis,n+1}[n+1] - k_{elong,n}[n][subunit] - k_{dis,n}[n], \tag{19.3}$$

where $k_{elong,n}$ and $k_{dis,n}$ are the rates of association and dissociation of subunits to a capsid of size n, respectively. The kinetics is split into a nucleation step, which involves the formation of capsid embryos, and an elongation process, which corresponds to the growth of the nuclei by the sequential addition of subunits. In this context, nuclei are treated as a special kind of intermediates that cannot disassemble and form at a slower rate k_{nuc} than the elongation stage, in order to justify the initial lag time of viral assembly. Numerical simulations have shown that this set of reactions quickly reaches a steady-state concentration of intermediates that asymptotically approach an apparent equilibrium characterized by a law of mass action, Eq. (19.1), from which one can obtain the association free energy of the capsid.

Using this model, it has been possible to describe the kinetics of *in vitro* assembly experiments of empty capsids of viruses like HBV [33] and HPV [36]. Further refinements were proposed by Dragnea et al. [37] to describe the assembly of BMV capsids, which include certain simplifications in the rate equations and an additional activation step for the monomers to fit the short time initial takeoff of the assembly. Remarkably, these studies conclude that the interactions driving the assembly are on the order of a few k_BT per protein, typically 5–6 k_BT per contact, which is much weaker than a covalent bond [33]. This has two clear biological motivations: first, weak interactions facilitate the correction of possible mistakes during the assembly; second, disassembly has to be feasible for many viruses to release the genetic material and infect a new host. Interestingly, the capsid of viruses like bacteriophages that do not require disassembling undergo a maturation process that in many cases reinforces the shell with additional or alternative subunit-subunit interactions (see Sect. 19.4.2 and Chap. 13).

However, despite the utility of these kinetic models, they present several important drawbacks. In particular, they involve a large set of reaction equations that need to be solved numerically, which can be very demanding computationally; the reaction rates must be fitted to the experimental data, which limits the predictive power of the theory; and the size of the nucleus is usually guessed a priori, which means that different cases should be explored for each experiment. Moreover, these models assume that the assembly is an equilibrium aggregation; hence, to justify that disassembly takes place at much lower protein concentrations than expected, they have to assume that the binding energies for assembly and disassembly are different.

Classical Nucleation Theory of Capsids

Alternatively, a recent approach based on classical nucleation theory (CNT), has overcome some of these limitations, explaining in a simpler way the common trends found experimentally in the assembly of capsids, and providing also a more direct connection between theory and experiments [20, 38]. The key point in CNT is the development of a physical model for the free energy of formation of partial capsids that, contrarily to the equilibrium aggregation theory discussed

above, contains explicitly the dependences in both the interaction between subunits and the concentration of free subunits. As we show below, this energy can be related to the rate of capsid production, reducing considerably the number of parameters to be fit, if any. Let's describe the main elements of CNT applied to viral capsids.

As we discussed above, the aggregation of a capsid becomes thermodynamically favorable when a building block in a fully formed capsid has a lower chemical potential than in solution, *i.e.*, when the interaction energy of a building block in the capsid overcomes the entropic penalty of removing it from the solution. Nevertheless, if one takes a closer look to the assembly process, one realizes that, in the successive addition of building blocks to intermediate structures, those CBBs placed at the rim of a partial shell will be exposed and miss few energy contacts (see Fig. 19.4), *i.e.*, the formation of intermediates intrinsically entails an energetic cost that originates an energy barrier. It is precisely this barrier what justifies the existence of a lag time at the beginning of the assembly process and the scarcity of intermediates observed experimentally. This is a well-known scenario in physics and is analogous to a vapor-liquid phase transition, where the formation of a liquid drop in the vapor phase has an energetic penalty associated to the surface tension of the drop. Following this analogy, the simplest way to model the energetic cost of an intermediate shell is by introducing a rim energy penalty associated to a line tension, equivalent to the surface tension in a drop.

In the framework of CNT applied to quasispherical capsids, a partial shell can be approximated as a spherical cap made of n subunits, whose free energy of formation is given by $\Delta G(n) = n\Delta\mu + a\sqrt{n(q-n)}$. Here $\Delta\mu = -k_B T \ln\left(\frac{c}{c^*}\right)$ is the difference in chemical potentials between a subunit in a complete capsid and in solution, and a is a constant related to the line tension. Even at favorable assembly conditions, the competition between these two terms gives rise to an activation barrier (see Fig. 19.6a). The maximum of this barrier defines the critical nucleus size

$$n^* = \frac{q}{2}\left(1 + \frac{\Delta\mu}{\sqrt{\Delta\mu^2 + a^2}}\right), \tag{19.4}$$

and its height

$$\Delta G^*_{as} = \frac{q}{2}\left(\Delta\mu + \sqrt{\Delta\mu^2 + a^2}\right) \tag{19.5}$$

defines the activation (or nucleation) barrier. As we see in Fig. 19.6a, for $n < n^*$ the free energy of formation of an intermediate increases with the number of CBBs, so these intermediates will tend to disassemble back into free subunits. On the other hand, partial capsids of size $n > n^*$ reduce their free energy upon the addition of CBBs, and they will tend to grow spontaneously until complete a capsid. Thus n^* is the critical size that partial capsids have to reach to trigger the formation of complete shells and can be considered as the *nucleus* or the embryo of the assembly

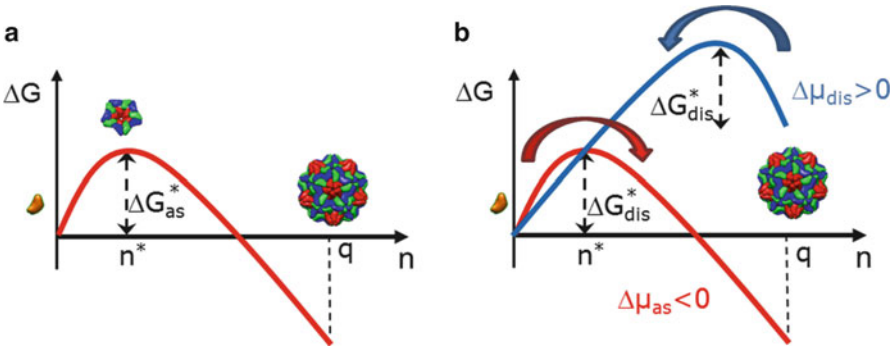


Fig. 19.6 Free-energy barriers of capsid formation in the context of CNT. (**a**) Total free energy of formation of a capsid as a function of the number of capsid building blocks n at favorable assembly conditions ($\Delta\mu_{as} < 0$). The complete capsid contains q building blocks. The nucleation barrier ΔG^*_{as} and the critical cluster size n^* are indicated. (**b**) Free energy of capsid formation for favorable assembly ($\Delta\mu_{as} < 0$, *red line*) and disassembly ($\Delta\mu_{dis} > 0$, *blue line*) conditions. For disassembly, the barrier has to be jumped in the opposite direction (*i.e.*, from *right* to *left*) than for the assembly. To have similar barrier heights for assembly and disassembly, different chemical potentials, and accordingly concentrations of capsid proteins, are required. That is ultimately the reason for the hysteresis between assembly and disassembly conditions found experimentally (see Fig. 19.5a)

process. It is worth emphasizing that, contrary to the assumption in the previous kinetic model, the critical nucleus is the most unstable intermediate because it corresponds to a maximum of the energy, and both its size and the height of the nucleation barrier depend on the concentration of subunits. In particular, for $c < c^*$ the formation of a capsid is not possible, and protein concentrations $c >> c^*$ are indeed needed to lower the nucleation barrier and the critical size enough to make the capsid formation feasible.

The rate of capsid formation J, *i.e.*, the number of capsids that form per unit volume and unit time, is controlled by the presence of this barrier. Like in other activation processes, it follows an Arrenhius behavior given by

$$J = c_s \beta^* Z e^{-\frac{\Delta G^*_{as}}{k_B T}}. \tag{19.6}$$

Hence it depends exponentially on the barrier height and linearly on β^*, which is the rate of attachment of CBBs to the critical cluster size. Here, Z is a correction factor associated with the local curvature at the top of the barrier and accounts for the possibility that nearly critical-sized clusters dissociate.

The activated mechanism described by CNT has several biological advantages. First, the requirement of a threshold concentration and the existence of a critical size that has to be reached to trigger capsid formation favors the production of enough proteins to guarantee the completion of capsids. In other words, if there were no barrier there would be multiple intermediates growing at the same time, which would rapidly deplete the protein by generating many incomplete shells, a

remarkably inefficient scenario. Additionally, the presence of a barrier is the origin of the hysteresis, which prevents the disassembly of capsids under conditions where the concentration of subunits is reduced or nonexistent, favoring the stability of the virus during the extracellular process before a new infection (see Sect. 19.5.2). A more detailed derivation of CNT for viral capsids can be found in Refs. [20, 38].

Simulations of the Assembly of Empty Capsids

Despite the advances in computational power, full-atom simulations of the assembly of even the smallest viruses are still out of reach. Accordingly, current simulations necessarily involve coarse-graining at different levels, which, apart from the computational advantage, represent also a great strategy to isolate the relevant details involved in the assembly process at a microscopic level.

At the coarser level, several models have explored the geometrical particularities required for the formation of T-number icosahedral shells. In particular, models based on local rules [39], which warrant the formation of the proper contacts between subunits, have shown that, due to the similarity between quasiequivalent positions (see Chap. 2), the set of rules to successfully assemble a T-number capsid can be less than T. For instance, four local rules suffices to produce $T = 7$ shells [40]. In a similar spirit, tiling approaches have studied the assembly of capsids focusing on the combinatorial aspects involved in the addition of subunits to intermediates of all possible sizes, concluding that only a small set of intermediates are relevant for the assembly pathway of empty shells [41, 42]. Various models have introduced a more physical implementation of these ideas by proposing effective interaction potentials with local constraints or strict orientational forces designed to produce a particular T-number target. This embraces patchy particle potentials [43] and bond-based potentials [44] (see Fig. 19.1).

The next level of complexity for coarse-grained strategies includes models that aspire to partially represent the shape of the assembly building blocks, like protein monomers, dimers, or pentamers and hexamers with a different degree of accuracy (see Fig. 19.1). In this way, subunits have been represented as pyramids [8], ellipsoids [9], triangular, and trapezoidal elements [10, 12, 45], which are agglomerates of relatively small spherical beads that help to place the interactions sites that are handcrafted to succeed in the assembly of a specific target structure.

Models based on a continuum elastic description of the assembly, based on growing a triangulated network, have also been successfully used to describe, for instance, the growth of spherical and polyhedral shells [46], and the (irregular) shapes of conical virus capsids like HIV [47, 48].

Finally, higher resolution coarse-grained strategies have been more challenging to investigate. However, some groups have studied, for instance, the assembly of HIV capsid proteins [24].

These simulation studies have confirmed many of the assumptions discussed in the aggregation and CNT theories, and have provided a better understanding of the difficulties and ingredients required for the successful assembly of empty viral

capsids. The combination of these theoretical and simulation studies is opening the door to a higher understanding of the assembly of viral capsids, and hopefully will facilitate the development of novel antiviral strategies.

19.3.2 RNA Co-assembly

Although *in vitro* experiments have demonstrated that, in certain conditions, the viral capsid proteins of ssRNA viruses alone are able to reconstruct the native capsid structure, *in vivo* their assembly generally requires the presence of the genetic material [49] (see Chap. 12). Consequently, the self-repulsion and entropy of the nucleic acid, and its interaction with the inner surface of the capsid are new ingredients that have to be considered to characterize the free energy of the complete system.

In physiological conditions ssRNA behaves like a flexible polymer with a persistence length on the range of the nucleotide size. In this scenario the main driving force in the co-assembly process is the electrostatic interaction between the ssRNA and the positively charged residues of the capsid, which often concentrate in disordered terminal segments of the capsid proteins. By virtue of electroneutrality one would expect a 1:1 correspondence between the length of the viral RNA and the charge of the inner part of the capsid in contact with the genetic material. However, experiments have determined that a variable part of the nucleic acid negative charge is neutralized by metal ions and/or polycations (*e.g.*, spermine or spermidine) in different viruses. Moreover, the distribution of charges in the capsid, the excluded volume effects and the Donnan potential (*i.e.* the difference in electrostatic potential between inside and outside the capsid due to the unbalance of charged macromolecules) have been shown to play a crucial role in the possible overcharging of viruses. Accordingly, there seems to be no universal genome to capsid charge ratio [50].

Several models have investigated the thermodynamics of co-assembly combining basic electrostatic and polymer physics theories, as discussed below. As a starting point, the complexation energy between ssRNA and the capsid can be evaluated using the DH mean field theory [32]. For spherical thin shells this yields $\Delta F_c = \frac{\pi\sigma^2 R^2}{\varepsilon_0 \varepsilon_r}(fd - \kappa^{-1})$, where d is the thickness of the ssRNA layer adsorbed on the inner capsid wall, f is a numerical factor, and κ^{-1} is the Debye screening length. The co-assembly is thermodynamically favored when this complexation energy becomes negative, which depends on the competition between the characteristic scales d and κ^{-1}. Contrary to the case of dsDNA bacteriophages discussed below, here the nucleocapsid reaches an energy minimum that can be evaluated from the previous energy of formation, indicating that ssRNA viruses may not be pressurized [32, 51]. More refined thermodynamic models that include the distribution of charges in the capsid, the possible presence of charged peptide arms, and the contribution of charged macromolecules, have been developed and can justify the

overcharging of some ssRNA viruses in nature and the favorable co-assembly of viruses with no net capsid charge [50].

Remarkably, many RNA viruses store several strands of genetic material rather than a long single sequence. This was thermodynamically characterized using a simple Flory mean-field theory [52], which accounts for the elastic compression of the chain, the self-repulsion of the genetic material, and the interaction of the polyelectrolyte with the inner wall of the capsid. This theory also justified the results from several *in vitro* experiments showing that the genome length can control the size of the final capsid [53]. Interestingly, the results of these thermodynamic models are valid also for other negatively-charged polymers different from RNA, as has been proved experimentally [28].

Regarding the organization of ssRNA, experimental studies have shown that it is distributed inhomogeneously inside the capsid, being very dense close to the inner surface and looser as we move towards the center, which in many cases is completely devoid of genetic material. The RNA in the central region does not adopt any particular structure and its organization is a consequence of the self-repulsion of the genetic material [54]; instead for some viruses the outer part sews the capsid proteins tracing a Hamiltonian path (*i.e.* a path that visits each vertex of the capsid only once) through the edges of a dodecahedron partially recovering the icosahedral symmetry of the shell, which could have an important role during the assembly of the capsid [26, 49].

Recently, several simulations have started to investigate the co-assembly of RNA viruses, but again this is a challenging task that requires coarse-graining and strong computational efforts. One of the first models explored the encapsulation of a polymer defined on a cubic lattice, but more realistic off-lattice simulations with flexible polymers encapsulated by coarse-grained capsomers or capsid proteins have been feasible lately [11, 55]. These simulations have assessed the role played by the genome length, the relative stoichiometry of genome *versus* capsid proteins, and the strength of the different interactions in the efficiency of the encapsulation process. In particular, it has been found that there is an optimal genome length that maximizes the encapsulation efficiency, which depends on the competition between capsid-capsid and capsid-genome interactions. Remarkably, depending on these interactions, two plausible assembly pathways have been identified (see Fig. 19.7). In the first one, the polymer is encapsulated in concert with the capsid formation, a process that is the natural extension of the assembly of empty capsids. Interestingly, in this case the capsid formation can be also described by classical nucleation theory, where the genome reduces significantly the nucleation barrier [55]. The other pathway takes place for strong subunit-RNA interactions. Initially many subunits (typically in excess to the total needed) attach to the genome and then they cooperatively rearrange to form the closed capsid. However, the process of nucleic acid-assisted assembly of ssRNA viruses is still far from well understood; further experimental and theoretical studies are clearly required, and the interested reader is referred to Chap. 12 for current views on the process based on experimental observations.

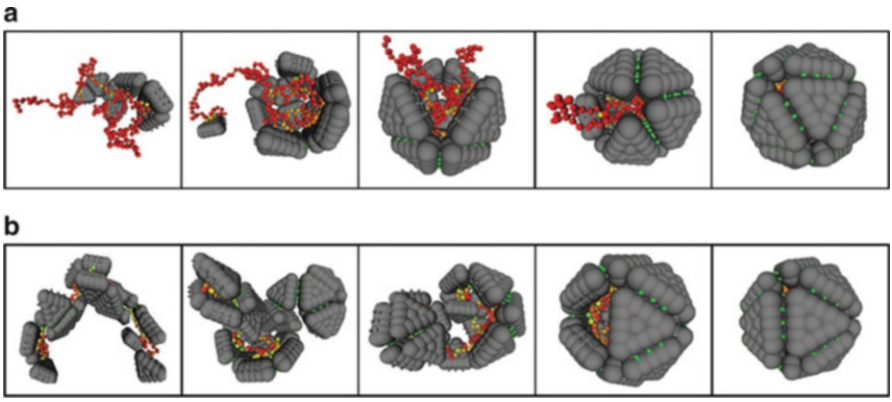


Fig. 19.7 Mechanisms of co-assembly for RNA virus. Different snapshots of a coarse-grained simulation showing two possible pathways in the formation of an RNA virus particle. (**a**) The RNA (*red chain*) is encapsulated in concert with the formation of the capsid by the addition of capsid proteins (modeled by multiple *grey* beads with *green* interaction sites). (**b**) For strong RNA-capsid protein interactions, first, many capsid subunits (in excess above the stoichiometric ratio) attach to the RNA in a disordered way and then cooperatively rearrange to form the complete capsid (Reproduced with permission from Ref. [11])

The encapsulation of non-genetic materials, such as gold nanoparticles, emulsions or negatively-charged polyelectrolytes by viral proteins, has also been explored in experiments and simulations, mainly driven by their potential technological applications (see Chap. 22).

19.3.3 dsDNA Packaging

There are two physical properties of dsDNA that play a crucial role in the assembly of many viruses: elasticity and charge. Contrary to ssRNA and ssDNA, dsDNA is a relatively stiff polymer. In a first approximation, its rigidity is characterized by its persistence length, *i.e.*, the length over which the DNA keeps straight under thermal fluctuations, which is about *50* nm for typical monovalent salt physiological conditions. This persistence length is similar to the characteristic capsid dimensions, and since dsDNA is usually stored at high densities, this implies that a significant elastic cost is paid during the packaging of the dsDNA genome inside the protective viral shell. Regarding the dsDNA charge, every base pair contributes with two negative charges. As for the viral ssRNA and ssDNA genomes, part of the dsDNA charge may be neutralized by metal ions and/or polycations. But in general, the charge causes the DNA chains to repel each other. The electrostatic term tends to dominate over the elastic one, although the latter seems to determine the conformation adopted by the confined dsDNA [56]. These two factors do not facilitate a spontaneous co-assembly of the capsid, except in the presence of condensing proteins like in

polyomavirus or adenovirus, and this is, most likely, one of the main reasons why many dsDNA viruses first assemble an empty capsid, the procapsid, and then package the genetic material using a molecular motor (see Chap. 12).

Along the years, several models have evaluated the energetics of the packed DNA [57–59]. Here we will describe the so-called inverse spool model that assumes that dsDNA is wrapped around with local hexagonal order in concentric hoops packed from the outside of the capsid towards the center (see Fig. 19.8). This theoretical approach assumes that the total free energy of confined DNA consists of the two contributions, elastic and electrostatic, discussed above. Then the total energy is given by $G_{tot}(L,d_s) = G_b + G_{int}$ and depends on both the total length of the genetic material L and the separation between strands d_s. The elastic contribution, G_b, is calculated assuming an inverse spool arrangement of the DNA, where the strands are packed in a hexagonal array with a spacing d_s. Then the elastic energy and the total DNA length become

$$G_b(L) = \frac{2\pi\xi_p k_B T}{\left(\sqrt{3}d_s\right)} \int_R^{R_{out}} (N(R')/R')dR'. \tag{19.7}$$

$$L = \frac{4\pi}{\left(\sqrt{3}d_s\right)} \int_R^{R_{out}} R'N(R')dR'. \tag{19.8}$$

where ξ_p is the persistence length, $N(R')$ is the number of hoops of radius R' in the capsid, and R and R_{out} are the inner and the outer radius of the inverse spool, the latter taken as the radius of the inner surface of the capsid (see Fig. 19.8).

An accurate first principle calculation of the DNA-DNA interactions is a difficult and unachieved task. Thereby, the electrostatic interaction G_{int} has been commonly characterized phenomenologically from osmotic stress experiments at the proper conditions [60]. These experiments have shown that the electrostatic interaction between DNA strands has a different behavior depending on the valence of the added salt. For monovalent and divalent salts, the contribution is purely repulsive and the dependence of the osmotic pressure $\Pi(d_s)$ with respect to the separation between strands is given by

$$\Pi(d_s) = F_0 e^{-\frac{d_s}{c}}, \tag{19.9}$$

where c and F_0 are parameters that characterize the decay length and strength of interactions, which depend on salt conditions. In the case of trivalent and tetravalent salts the effective electrostatic interaction has an optimal distance d_0 between strands, *i.e.*, at smaller separations leads to a repulsive interaction whereas at higher separations promotes attraction between the strands. A convenient empirical expression for the osmotic pressure in this situation is

$$\Pi(d_s) = F_0\left(e^{-\frac{(d_s-d_0)}{c}} - 1\right). \tag{19.10}$$

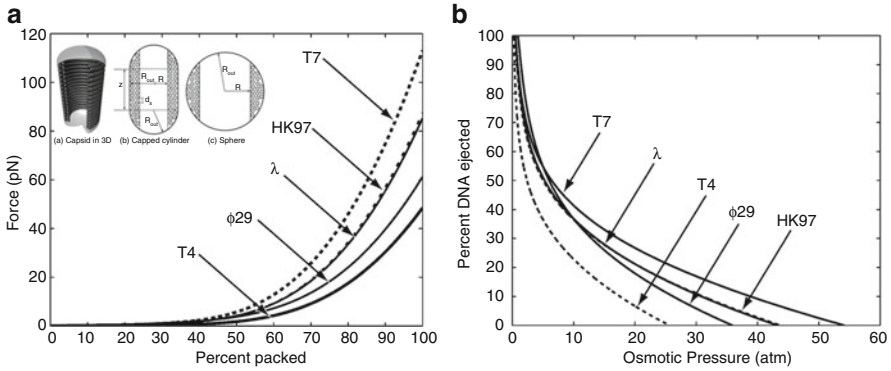


Fig. 19.8 Forces and pressures of confined dsDNA. (**a**) Forces resisting dsDNA packaging (and assisting dsDNA ejection) estimated using the inverse spool theory for different bacteriophages (T7, λ, T4, Φ29, HK97) in 0.5× PBS buffer. The inset shows the idealized organization of the dsDNA inside different capsids assumed in the model. (**b**) Osmotic pressure inside the capsid (in the horizontal axis) as a function of the percentage of DNA ejected estimated by the inverse spool theory, for different bacteriophages (T7, λ, T4, Φ29, HK97) in a buffer containing10 mM $MgSO_4$ (Figures reproduced with permission from Ref. [58])

From the osmotic pressure, the interaction energy can be obtained yielding

$$G_{\text{int}} = \sqrt{3}F_0 L\left(c^2 + cd_s\right)e^{-\frac{d_s}{c}} \tag{19.11}$$

in the purely repulsive regime and

$$G_{\text{int}} = \sqrt{3}F_0 L\left[\left(c^2 + cd_s\right)e^{-\frac{(d_s - d_0)}{c}} - \left(c^2 + cd_0\right) - \frac{1}{2}\left(d_0^2 - d_s^2\right)\right] \tag{19.12}$$

in the presence of tri- and tetra-valent counterions. It is important to emphasize that this interaction free energies G_{int} obtained from osmotic stress experiments not only account for the electrostatics of the DNA and counterions, but also for entropic and hydration contributions [58, 60].

Knowing the contributions of the bending and interaction terms, the optimal energy and DNA spacing are then obtained by minimizing the total energy $G_{tot}(L,d_s)$ with respect to d_s for a given total genome length L. The internal pressure P of the confined DNA or the force F at which it will be ejected can be finally obtained from it by their standard thermodynamic definition as $P = -\partial G_{tot}/\partial V$ or $F = -\partial G_{tot}/\partial L$. This model predicts the forces during the ejection process and the internal pressure in bacteriophages [58, 61], which are in the range of 20–60 atmospheres as determined by AFM and osmotic pressure experiments [62, 63] (see Chaps. 9 and 18 and Fig. 19.8). Remarkably, bacteriophages require the strongest molecular motors reported so far to store the dsDNA at such high densities. In many cases this stored energy is used to initially drive the injection of the dsDNA genome into a bacteria during the infection process.

Although the fact that the dsDNA is compacted in an inverse spool configuration is still controversial [56], the estimates of the pressures are not very sensitive to the configuration, since they are mostly controlled by the electrostatic contributions [64] and the packing density of the DNA. In fact, using in Eqs. 19.9 and 19.10 a simple estimate of the strand separation

$$d_s = \sqrt{\frac{2V}{\sqrt{3}}}\frac{1}{\sqrt{L}}, \tag{19.13}$$

which is obtained by equating the total volume available inside the capsid V with that of the closed packed DNA (assuming that it occupies the whole volume), one gets an accurate approximation of the capsid pressure due to the confined DNA.

19.4 Mechanical Stability of Capsids

Several studies indicate that virions can be subjected to substantial mechanical stress, for instance, due to osmotic shock, related to a sudden change of salt concentration in the environment, or to the packaging of the viral genome at high densities (see Chaps. 12 and 18). Interestingly, the last scenario concerns many dsDNA bacteriophages, which have to withstand up to tens of atmospheres arising from the confined viral genome, as we have just seen. Additionally, after the self-assembly of the capsid, to become infective many viruses undergo a maturation process that may properly tune their physical and mechanical properties. In bacteriophages, this maturation often involves the modification of the capsid shape through the so-called *buckling transition*, where an initial spherical procapsid undergoes a transformation into a polyhedral-like shell resembling an icosahedron (see Chap. 13). All these complex scenarios illustrate different sophisticated strategies that viruses have developed to protect their genetic material, remaining infective despite the hazards of the environment. Thus, their mechanical characterization becomes of particular interest in order to understand the stability of viruses.

Very little is known yet on the mechanical properties of viral capsids. A few studies relied in *bulk* experiments using biochemical and physical techniques, like inelastic Brillouin light scattering. However, the recent application of atomic force microscopy (AFM; Chap. 8) and optical tweezers (Chap. 9) have allowed the experimental determination of the mechanical properties and mechanochemical function of individual viral particles (see Chaps. 8, 9 and 18). This has also boosted the interest in developing and applying theoretical frameworks to characterize the mechanical response of viral capsids.

Thus, in this section we will discuss the main theoretical approaches applied for the mechanical characterization of viral capsids. The studies are grouped in three

blocks: first, we introduce some theoretical strategies to investigate the elastic response of viral capsids, which are essentially based on the TST framework and computational simulations; then some physical ideas behind the maturation of viruses are discussed; finally, we briefly review some techniques that, by analyzing thermal vibrations of the capsid, facilitate the extraction of important structural properties and general information about shell transformations.

19.4.1 Elastic Response of Viral Capsids

Elastic Properties from TST

TST constitutes the natural starting framework to study the mechanical properties of viral capsids. This model has been particularly useful to complement AFM indentation experiments (see Chap. 18), facilitating their interpretation and allowing the extraction of the elastic properties of the capsid [65]. To describe the local deformations of the capsids generated in AFM experiments, it is more convenient to rewrite the elastic energy in terms of a free energy functional of the indentation profile $\zeta(\mathbf{r})$ given by $\Delta F = \int dS \left\{ \frac{1}{2}\kappa(\Delta\zeta)^2 + \frac{1}{2}Y\left(\frac{2\zeta}{R}\right)^2 + \frac{1}{2}\tau(\Delta\zeta)^2 \right\}$. The first term describes the bending energy (with $\Delta\zeta$ being the curvature), the second term the stretching energy, and the third term the extra work required to deform the capsid, assuming that the shell is subjected to a tension τ [65]. The indentation profile can be obtained from the differential equation $\delta\Delta F/\delta\zeta(\mathbf{r}) = f(\mathbf{r})$, where $f(\mathbf{r})$ is the applied radial force per unit area. The previous equation is valid only for small deformations and can be solved analytically only in special situations, for instance, in the case of a point force applied on an empty spherical shell. Let's focus first on the scenario where the tension τ, which can be related, for example, to an internal pressure, vanishes.

For indentations much smaller than the shell thickness, TST predicts a linear relation between the applied force and the resulting indentation, as follows from the simple scaling arguments described in Chap. 18. The corresponding effective spring constants (as a measurement of mechanical stiffness, see Chap. 18) for spherical and cylindrical shells are $k_{sph} = \frac{2}{\sqrt{3(1-\nu^2)}}\frac{Eh^2}{R}$, and $k_{cyl} \propto Eh\left(\frac{h}{R}\right)^{3/2}$, respectively [17], where E is the (three-dimensional) Young's modulus, ν is the Poisson ratio, R is the radius of the capsid, and h its thickness. As shown in Chap. 18, using these expressions the Young's modulus of a virus can be derived from the slope of the force indentation curves obtained with the AFM, provided the size and the thickness of the capsid are known.

In addition, as described in Chaps. 9, 18 and Sect. 19.3.3, the presence of the confined dsDNA genome can build up a substantial pressure in the capsid, which may ultimately modify its mechanical response. In a first approximation, the classical Laplace law is able to estimate the tension generated in a shell by a

pressure difference Δp, which for a spherical shell is $\tau = \Delta pR/2$. Using this value in the previous free energy functional, and solving for a point force, one obtains the effective spring constant [66],

$$k = k_{empty} \frac{\pi}{2} \frac{\sqrt{\tilde{\tau}^2 - 1}}{\arctan \mathrm{h} \sqrt{\left(1 - 1/\tilde{\tau}^2\right)}}, \tag{19.14}$$

which depends on the ratio $\tilde{\tau} = \Delta pR/k_{empty}$ between the pressure difference Δp multiplied by the radius R and the spring constant k_{empty} of the empty shell. This expression provides a simple way to estimate the internal pressure of spherical viruses by comparing the spring constants of the emptied mature capsid and the virion, provided that the cause of this difference is indeed the existence of internal pressure (see Chap. 18).

It is worth noting that experiments and calculations have shown that tension does not necessarily originate from an internal pressure. For instance, in the icosahedral minute virus of mice, viral ssDNA segments bind to equivalent sites at the capsid inner wall and act like molecular buttresses that reinforce anisotropically the viral particle (see Chap. 18); in the case of bacteriophage Φ29, it has been shown that the capsid is also mechanically anisotropic because of built-in stress [67]. The development of pre-stress is a common strategy in engineered structures and probably reinforces the capsid in order to better tolerate the pressure developed during the packaging of the viral genome.

An important mechanical property of a capsid, especially for dsDNA viruses, is its resistance to pressure before bursting. TST provides simple expressions that relate the largest pressure that a capsid can resist with the maximum relative radial expansion and the maximum tangential stresses. The largest stresses can be in turn evaluated in terms of the inflection point of the potential between subunits, *i.e.*, the maximum force that the interaction can hold [68, 69]. In addition, by simply using the Laplace law, one obtains that the maximum pressure Δp^* that a virus can tolerate, at equal interaction strengths, decays with the radius of the shell as $1/R$. In other words, at equivalent values of the molecular interactions, big capsids should be less efficient tolerating pressures. Notice that this decay also applies for the effective spring constant k, and it can be also shown that this follows for the bulk modulus [20, 70].

Finally, TST and scaling arguments can also be used to characterize large indentations of viral capsids using point-like axial loads. For large deformations the response of the viral capsid becomes strongly non-linear, which in many cases is associated to an inversion of the curvature (*buckle*) in a region of the shell around the indentation point [17, 71]. The elastic analysis shows that for spherical shells this inversion typically takes place when the indentation becomes comparable to the shell thickness h, or alternatively, when the load force exceeds a threshold value $F_{inv} \sim \kappa/R$. In the nonlinear deformation regime, the force-indentation curve is expected to show a squared root dependence on the indentation [17, 71]. However,

in these situations the characterization of the indentation using TST becomes very challenging, and other techniques, like finite element analysis (FEA) discussed in Chap. 18, are more suitable. FEA studies have demonstrated that capsids become stiffer and show higher non-linearities (associated to local bulging or buckling) for high FvK numbers [19, 65]. The relevance of the finite thickness and non-uniformity of viral shells has been also explored, for instance, in the softening of CCMV after its pH swelling transition, or in the anisotropic response of HBV when indented at the different axis of icosahedral symmetry [19, 65].

Discrete Models and Simulations

However, FEA is essentially a numerical solution of continuum elasticity equations, and accordingly it shares also certain limitations. For instance, the plasticity or the breaking of viral shells observed after repeated indentations cannot be captured. In addition, the discrete and possible anisotropic interactions between capsid building blocks cannot be easily incorporated in this description. To overcome these limitations, one has to resort to the use of discrete models.

Unfortunately, contrary to the case of the structure and assembly of viral capsids (described in Sects. 19.2 and 19.3), there have not been yet many models of this kind studying the deformation of viral capsids. The most popular simulation methods have been based on the discrete triangularization of TST, described in Sect. 19.1 [72], which have been able to reproduce the non-linearities in the indentation of viral capsids and estimate mechanical differences between some T-numbers. It is worth noticing that this should not be confused with FEA, which discretizes the capsid in tiny elements that do not correlate with the positioning of the capsid proteins in the shell.

Also at the low-resolution level, the model of Ref. [69] has highlighted the importance of the particular arrangement of the capsomers and the triangulation number on the mechanical properties of spherical capsids. At the high-resolution level, Schulten's group has recently developed a coarse-grained model of HBV complementing AFM experiments. The simulations were in agreement with the experiments without the necessity of any type of fitting [65, 73]. One of the most interesting features of this coarse-grained approach was the identification of the local rearrangements of capsid proteins and the irreversible conformational changes produced during the indentation of HBV [74], which was not captured by the FEA studies of HBV.

19.4.2 Maturation and Buckling

As described in Chap. 13, to become infective many viruses require a maturation step after the self-assembly of the procapsid. This process seems to tune different properties of the viral shell that are manifested in many cases by a noticeable

change in the capsid shape, generally associated to a *buckling transition*, which is of particular relevance among bacteriophages. In this transformation, the initial spherical procapsid with icosahedral symmetry undergoes a transition into a polyhedral shell with flatter triangular faces and an overall shape resembling an icosahedron.

TST provides again an excellent framework to establish the physical basis of the buckling transition [4]. In this context the asphericity of the capsid, *i.e.*, the degree of buckling, is exclusively determined by the FvK number, $\gamma = YR^2/\kappa$, which compares the stretching and bending contributions in the shell. When FvK exceeds a certain threshold value, $\gamma_B \sim 10^2$, the stretching energy of a spherical shell can be reduced if the 12 pentamers buckle out inducing the faceting of the shell into a polyhedral shape (see Fig. 19.2). Remarkably, it can be shown that within TST, the FvK number can be expressed as $\gamma = 12(1-\nu^2)(R/h)^2$, which means that it essentially depends on the square of the ratio between the capsid radius R and its thickness h. This suggests that big and thin shells should be more likely to adopt polyhedral rather than spherical shapes, in agreement with experimental observations of the faceted nature of large viruses [70].

Alternatively, a low-resolution discrete model has shown that the buckling transition depends on the triangulation number T and the icosahedral class[2] P of the virus structure [70]. In particular, P = 1 shells are most likely to produce polyhedral shapes, whereas viruses with P = 3 prefer to remain spherical (see Fig. 19.9). For big capsids the polyhedral shape becomes systematically the most stable, in agreement with TST. This study showed that in general the buckling transition can be also induced by a capsid expansion, in consonance with virus maturation. Furthermore, in comparison with the spherical shape, the resulting icosahedral shell is mechanically stiffer, tolerates larger expansions, and withstands higher internal pressures before failing. This suggests that the polyhedral shape could have a certain advantage for dsDNA phages that rely on the pressurization of their genetic material to be infective, justifying their buckling transition.

However, despite the very useful insights provided by the theoretical tools described above, one should not forget that the occurrence of the buckling transition during virus maturation is a complex process that often involves different molecular mechanisms including conformational changes, cleavage, and even the formation of covalent bonds. These mechanisms are virus-specific and cannot be easily characterized by most of the models described previously.

[2] The icosahedral class P was introduced by Caspar and Klug to classify triangulation numbers and it is related to them by $T = h^2 + k^2 + hk = Pf^2$, where f is the greatest common divisor of the integers h and k. It should not be confused with the pseudo-triangulation number P, used to label icosahedral viruses when the subunits that occupy quasiequivalent positions are chemically or structurally different (see Chap. 2).

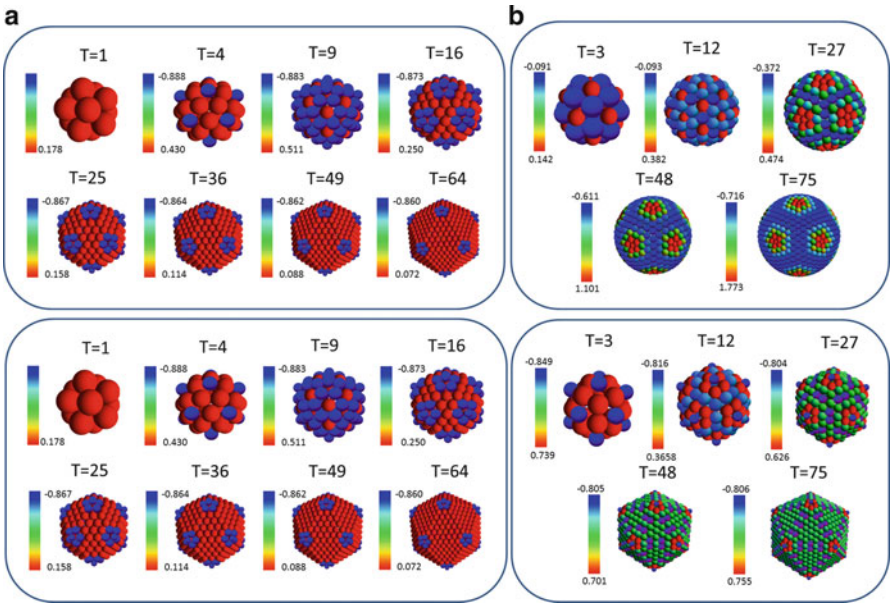


Fig. 19.9 Relevance of capsid structure in the buckling of spherical viruses. (**a**) The distribution of local stresses (indicated by the *color scales*) for spherical (*top panel*) and polyhedral (faceted, *bottom panel*) T-number shells belonging to the class P = 1. Capsomers in blue are stretched, whereas those in *red* are compressed. In the spherical capsids, positive stress concentrates on the lines connecting pentamers for T > 4, delimiting triangular areas with stretched hexamers. The absolute values of these stresses get larger as the T-number of the shell increases. The hexamer stretching is relieved if the triangular regions between pentamers are flattened; thus, these structures have a strong preference for buckling. (**b**) Same for spherical and polyhedral (faceted) T-number shells belonging to the class P = 3. In spherical P = 3 shells, compression concentrates on the pentamers, whereas the hexamers surrounding them are slightly stretched, leading to a dodecahedral pattern instead of the icosahedral one observed in P = 1. In this case, nothing indicates any particular relief of the stress on the hexamers upon adopting an icosahedral shape (Reproduced from Ref. [70])

19.4.3 Viral Capsid Vibrations

An interesting fundamental property of materials is the relation between the unavoidable thermal vibrations that they suffer at finite temperatures and their structural properties. In the case of capsids, this has been largely explored applying normal mode analysis (NMA) of elastic networks based on all-atom or cryo-EM structural models of specific viruses [75]. Essentially, in NMA the interaction between atoms or pseudo-atoms is approximated by a harmonic potential, and the individual vibrations of the subunits are characterized in terms of collective harmonic oscillations, called normal modes, which can be obtained by diagonalizing the second derivative of the potential energy (Hessian) [76]. In this description the

low frequency modes are associated with conformational changes in the capsid, and the high frequency modes describe localized motions.

In this way, NMA has allowed the study of the transition during the swelling of CCMV, and the maturation of *Nudaurelia capensis* ω virus and phage HK97 [76]. Recent investigations, combining high-resolution simulations and NMA, have shown that the maturation of HK97 follows a pathway dominated by icosahedral symmetry [77]. Remarkably, the combination of full-atom simulations and elastic network models constitutes a promising alternative method to obtain the Young's modulus of viruses that should complement the values obtained by the combination of TST and AFM [78]. Interestingly, these multiscale techniques have recently shown that cross-linking between capsomers during the maturation of HK97 is responsible for the increase of the shell resistance to breaking without noticeable changes in the Young's modulus [79].

19.5 Genome Delivery and Virus Egress

In this section we are going to discuss some of the physics behind particular mechanisms of genome delivery, and virus exit from and entry into cells. Once again, the relative simplicity and lack of an own metabolic machinery of viruses have led to biological solutions that strongly rely on basic physical processes. First, we describe, in physical terms, the injection of the genome of dsDNA bacteriophages into the host cell. Then, a mechanism of capsid disassembly is discussed in the context of CNT, which is important for many ssRNA viruses. Finally, we describe the modeling of the endocytosis and budding processes, which respectively mediate the entry and exit of many viral particles in the host.

19.5.1 Active and Passive Translocation

Bacteriophage infection relies on the translocation of the viral genome into the bacteria, leaving an intact empty capsid in the extracellular environment. In this infection pathway, the genetic material must rapidly get inside the cell, and the efficiency of the process is accomplished by a combination of different mechanisms, some of them not fully characterized yet (see Chap. 17).

In order to succeed in this translocation pathway, many bacteriophages rely on a proteinaceous tail that is attached to the capsid (head) and helps to initiate and direct the injection of their genomes through the bacterial wall and membrane (see Chap. 17). This process can involve complex conformational changes in the tail proteins requiring energy. A typical example is the contractile tail of bacteriophage T4. The contraction of the tail sheath drives the rigid inner tube through the bacterial wall and allows the virus to inject its DNA into the bacteria (see Chap. 17).

This contraction has been described in the context of physics as a *martensitic transition* [80, 81], a well-known phase transition occurring typically in solids. The idea is that the tail proteins can adopt two possible conformational arrangements: one extended and strained, and the other contracted and relaxed. A small activation barrier separates both states, where the initial one (the extended) has a higher energy. When the tail attaches to the membrane it triggers a rapid transition that releases the elastic energy stored by the proteins in the extended conformation adopted during the assembly of the virus. The contraction force derived from these studies is about 100 pN, roughly twice the stalling force of the motor packaging the DNA of Φ29 (see Chap. 9). This type of martensitic mechanism seems to be also behind other large conformational changes in the fusion proteins of different enveloped viruses, such as HIV and influenza virus [80] (see Chap. 16).

After a triggering event the dsDNA starts to enter the cell. As mentioned in Sect. 19.3.3, the capsid internal pressure drives the initial ejection of the genetic material. The initial forces and velocities of this entrance can be properly quantified by the inverse spool model, but after roughly the 50 % entrance of the DNA the pressure difference vanishes, and the remaining genetic material relies on other mechanisms to get in (see Chap. 12).

Different physical strategies can be conceived to explain the entrance of the remaining viral genome length inside the cell. The simplest one is by pure diffusion. In this case, thermal fluctuations shake the nucleic acid back and forth through the cell wall pore until the genome completely enters into the host. However, the translocation time of this mechanism scales as the square of the genetic material length L leading to $\tau_{dif} = L^2/D$, where D is the diffusion coefficient of the chain, and for typical viral parameters this time becomes prohibitively long. Alternatively, a second mechanism involving the presence of binding proteins has been studied to explain how the remaining viral genome is driven into the cell. Here, as soon as a portion of DNA gets inside the cell, certain proteins bind to it and impede its diffusion backwards, which rectifies the diffusion of the chain and speeds up the entrance by a factor proportional to the number of binding sites. Nevertheless, despite this rectification, the entrance time estimated for this mechanism is still too long. These two simple strategies cannot explain the entrance of the viral genome. However, it was recently shown that the proteins that bind to the DNA not only rectify its diffusion but simultaneously exert an effective translocating force that considerably increases the speed of the DNA entrance [82]. This additional force was characterized by a simple model evaluating the free energy of a stiff portion of DNA entering a spherical cell in the presence of nonspecific binding proteins (see Fig. 19.10). This simple model shows that the total free energy becomes more favorable the more genetic material gets inside the cell translating into a powerful effective force that drives the translocation. The effective force observed in coarse-grained simulations was also quantified using this free energy in a simple non-equilibrium kinetic model. Interestingly, this speed up effect associated to the binding energy of proteins has been verified in *in vitro* experiments that measure the kinetics of ejection (in that case in the presence of DNAse) [83]. On the other hand, although the model can also be refined by accounting for the entropy of the

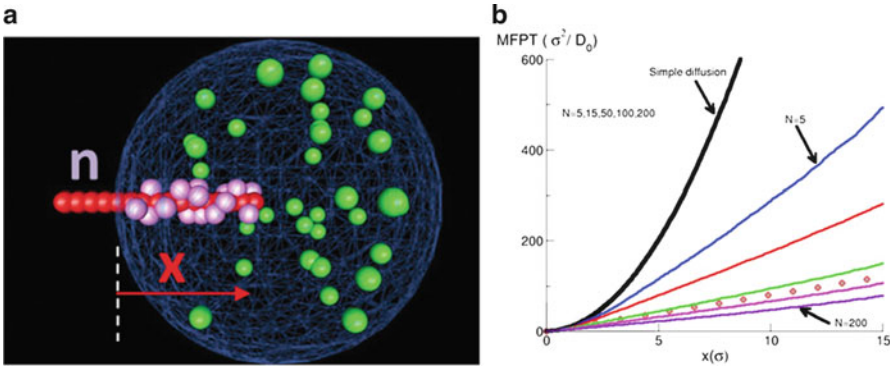


Fig. 19.10 Model for the translocation of a stiff portion of dsDNA in the presence of binding proteins. (**a**) Schematic illustration of the simulation setup, where a stiff portion of dsDNA (in *red*) enters a cell (in *blue*) in the presence of binding proteins (*green* (unbound) and *violet* (bound)). (**b**) Average time required for the entrance of a given distance x (measured in units of the bead size σ) by simple diffusion (*black line*) and in presence of different number N of binding proteins. For a sufficiently large concentration, the entrance becomes even faster than by ratcheted diffusion (*red dots*) (Reproduced from Ref. [82])

DNA chain [61], there are still many puzzling features of this translocation process of bacteriophages to be described, as has been shown by recent experimental studies [84].

DNA polymerases can also play an active role pulling in the DNA of some viruses, which has been studied experimentally with optical tweezers (see Chap. 9) and analyzed with stochastic models.

19.5.2 Capsid Disassembly

Eukaryotic viruses and some bacteriophages typically follow an infection pathway that requires virions to penetrate the host. Once inside, the capsid of these viruses have to change its conformation or, in many cases, disassemble to allow the release of the genetic material in specific regions of the cell, which will ultimately initiate the viral replication factory (see Chap. 14).

In many cases, viruses follow a post-assembly maturation stage that strengthens the initial weak interactions responsible for the assembly (see Chap. 13). Nevertheless, other viruses do not necessarily go through a maturation process, but they have still to avoid disassembly during their trip to infect a new host. If the assembly of viruses were an equilibrium process, viruses would disassemble as soon as the capsid protein concentration drops below the threshold value c^*. In this context the models introduced in Sect. 19.3.1 can be invoked. In particular, the CNT offers a good explanation of how viruses may accomplish this delicate balance. It is precisely the existence of an activation barrier that had to be overcome during the

assembly what prevents their disassembly. Then the process of viral disassembly can be though as the inverse of the assembly and can be mathematically described using the same CNT theory [20].

The disassembly of viruses like HBV has been probed by *in vitro* experiments (see Fig. 19.5a). This virus does not suffer any maturation or conformational change and the HBV capsid proteins can be reassembled repeatedly. In these experiments, one starts from a given total concentration of proteins in the form of complete capsids, and dilutes this concentration monitoring the disappearance of capsids. The results of these experiments have shown that the threshold concentration at which capsids disassemble, c_1^{dis}, is much lower than c_1^{as}, the value at which they start to assemble. This can be easily understood and quantified in the framework of CNT.

To initiate the disassembly, one has now to jump the barrier in the opposite direction respect the assembly (see Fig. 19.6b). The height of this disassembly barrier is given by $\Delta G^*_{dis} = \Delta G^*_{as} - \Delta G(q)$, where ΔG^*_{as} is given by Eq. (19.5). The disassembly rate depends also exponentially on ΔG^*_{dis} as in Eq. (19.6). At the conditions at which the assembly took place, this barrier is too high to be surmounted and capsids remain stable. In fact, positive $\Delta\mu = -k_B T \ln(c/c^*)$, implying c << c* are needed to lower the disassembly barrier up to a point that disaggregation is feasible. Since it is the height of the assembly and disassembly barriers what dominates the kinetics of both processes, a simple argument can be made to get a reasonable estimate of what would be the concentration required for disassembly. Assuming that the height of the barrier towards disassembly has to be the same than that of the assembly (to observe the processes at similar time scales) one gets the following simple expression

$$(c^*)^2 = c_1^{as} c_1^{dis} \tag{19.15}$$

that relates the threshold protein concentrations required for the assembly, c_1^{as}, and disassembly, c_1^{dis}, with the equilibrium concentration c^*, and through Eq. (19.1), with the binding energy per subunit. Thus, these simple ideas provide a way to extract useful information combining assembly and disassembly experiments [20].

19.5.3 Budding and Endocytosis

Many animal viruses require also crossing in and out cellular membranes to proceed with the infection. In enveloped viruses the lipid membrane and fusion proteins help to accomplish this task (see Chap. 16), and non-enveloped viruses are generally able to get inside the cell through endocytosis (see Chap. 15). Different physical factors control the feasibility of these processes: the membrane elasticity, the adhesion energy between the capsid and the membrane, the spontaneous thermal fluctuations, and the diffusion of spike proteins. Accordingly, several physical models have been proposed to address these interesting problems.

After the complete assembly of the nucleocapsid, most enveloped viruses acquire their envelope during the exit process from the cell by wrapping and pinching off of the host membrane. However, this deformation of the membrane entails an energetic cost. The first mechanism suggested to accomplish that was the spontaneous wrapping of the virus taking advantage of thermal fluctuations of the membrane [85]. However, the energy provided by thermal fluctuations alone is not enough to wrap completely the virus unless the membrane stiffness is low [86]. To overcome this cost, the virus counts on the adhesion energy between the membrane and its spike or capsid proteins. Several models accounting for these factors have been proposed to study the competence between these contributions and the optimal conditions for virus exit.

This process has been characterized by analyzing the wrapping of a spherical particle by an elastic membrane [87]. The elastic deformation of the membrane is described using a continuum elasticity model (see Sect. 19.2.2), where an additional term accounting for the adhesion energy between the virus and the membrane is introduced. This model has shown that a minimum value of the adhesion energy is required to have full wrapping, and that the process is very sensitive to the size of the viral particle. These results suggest the possibility of interfering with the entrance or budding of viruses by modifying the adhesion energy or the membrane tension. Recent simulations show that elasticity theory can account quantitatively for different mechanisms of budding, although some situations controlled by the kinetics of the process are not captured [88].

An extension of these ideas has been performed to account for the concurrent assembly and budding processes of HIV [89], ascertaining the conditions in terms of Gag-Gag interaction strength and membrane rigidity required to have complete budding.

During the budding and exit from an infected cell, many enveloped viruses embed in their lipid membrane spike-shaped proteins that play a crucial role to infect another cell. Statistical models of viral budding, which incorporate the entropy and distribution of spikes to the elastic and adhesion effective free energy terms, have also been proposed [90]. These studies analyzed the simultaneous formation of many buds and the competence of viruses to sequester the spike proteins. The results show that, for biologically reasonable values of the membrane elastic constant and viruses with relatively strong spike binding strength, all binding sites of the capsid get occupied by spikes upon successful budding.

Interestingly, with slight modifications, models developed to study budding can also be applied to study endocytosis. The endocytosis of the virus is mediated by the binding interactions between ligand molecules on the viral capsid and their receptor molecules on the cell membrane. These receptors are mobile and have to diffuse in the membrane to wrap around the virus and facilitate its entrance. In some case, this is the rate-limiting step for virus entry, and the corresponding kinetics can be modeled using a diffusion equation, as done in Ref. [91]. These studies show that particles in the size range of tens to hundreds of nanometers can enter (or exit) cells *via* wrapping without requiring the help of clathrin or caveolin. The optimal size to

get an optimal wrapping time was characterized, and turned out to be in the range of 25–30 nm in radius, which is quite similar to the size of many spherical viruses.

19.6 Perspectives and Conclusions

Theoretical models and simulations are starting to make a significant contribution to our understanding of how viruses function. As we have seen in this chapter, the regular architecture of viruses obeys general geometrical rules and basic physical principles. For instance, the icosahedral symmetry of spherical viruses and the structure of prolate viruses are a consequence of free energy minimization of generic interactions between the capsid proteins. In addition, the study of continuum elastic theories, and TST in particular, has established a reference framework to rationalize the distinguishing morphologies and the faceting of viral capsids, based on the minimization of the effective elastic and bending energies. Complementing TST, coarse-grained models have shown the dependence of buckling on the triangulation number, and high-resolution NMA have described basic aspects of capsid maturation for several viruses. The combination of TST, coarse-grained, FEA, and NMA simulations has been invaluable to characterize the remarkable elastic properties of capsids, facilitating the interpretation of AFM experiments.

In the context of viral assembly, CNT and kinetic models provide a solid background to explain the physical mechanisms involved in the assembly and disassembly of capsids. In CNT, the most important element is the existence of an energy barrier controlling the formation of capsids, which justifies the main features observed experimentally, *i.e.*, the scarcity of intermediates, the existence of a lag time at the beginning of the assembly, its sigmoidal kinetics, and the hysteresis in the disassembly. Coarse-grained simulations have started to complement CNT and other models by exploring the pathways of capsid assembly and the capsid-genome co-assembly of some ssRNA viruses, and they show that the genetic material is able to reduce the energy barrier of capsid formation. In addition, simulations have shown that the inhomogeneous distribution of RNA inside the capsid is a generic feature due to the self-repulsion of this highly flexible genetic material; several models have also justified the relation of charges between the genetic material and capsid proteins, the relation between the capsid size and the encapsulated cargo, the assembly competition of different ssRNA chains, or the ordered structure of the RNA layer in contact with the interior of the shell.

Alternatively, for some dsDNA viruses, it has been possible to estimate the pressures built up during the packaging of the genetic material and the corresponding ejection forces, which assist the initial stage of the dsDNA entry inside a new host cell. The translocation of the remaining material can be directed by either a nonspecific physical mechanism, like the binding of proteins to the viral genome, or by the pulling action of a polymerase. In the first case, the speed up due to the binding of proteins is associated to the lowering of the free energy with the

length inside and can be described by a simple adsorption model. Finally, statistical thermodynamical models are starting to describe the general physical mechanisms involved in the recruiting of the viral envelope and the membrane viral proteins, as well as the exit and entry of some viral particles by budding and endocytosis, respectively. The results indicate that viruses have the right size to efficiently favor the spontaneous occurrence of these processes.

Theoretical physical virology is a relatively young field, and most of the *smart* physical strategies developed by viruses to accomplish different stages in their life cycle are only starting to be unveiled. At the structural level, the understanding of irregular and conical capsids, and the mechanisms involved in the selection of a specific structure by a particular virus are major challenges to be faced. Regarding the mechanical stability of viruses, the main goal is to overcome the limitations of continuum techniques characterizing the elastic properties of viral shells. For this purpose, the development of coarse-grained discrete models, which incorporate the anisotropy of the interactions between the capsid building-blocks, is very timely and will facilitate a more accurate description of the protective role of the capsid, including the rich mechanical phenomenology obtained in AFM experiments. In the context of assembly, a clear direction is the further development of CNT to reach a quantitative and predictive understanding of assembly experiments on specific viruses, and its extension to explicitly incorporate the effects of the genetic material and auxiliary (scaffolding and/or condensing) proteins. Simulations will play a very important role to test these theories and to gain further insight into the molecular details of this process. Finally, regarding the delivery of the genetic material and the egress of viral particles, simulations and theory are needed to clarify specially the nonequilibrium and kinetic aspects of these processes.

Obviously, the answer to all these questions will require a close collaboration between physicists and molecular virologists. In this endeavor, the main challenge faced by theory is to properly adapt the main physical ideas into a precise biological context and relate them to accessible experimental magnitudes. This should allow us to progressively incorporate other important biological information, demonstrating the utility of theories in the prediction, guidance, and interpretation of experiments. With this effort, physics would provide not only a theoretical understanding of viruses but also a valuable tool in the development of novel antiviral strategies and useful applications.

Acknowledgements We acknowledge useful discussions with María Aznar, Robijn Bruinsma, Mauricio Comas-García, William Gelbart, William Klug, Charles Knobler, Joseph Rudnick, Jan Wedekind, and Roya Zandi. We also thank Charles Knobler for a careful reading of the manuscript and Alexander Bittner, and Carmen San Martín for fruitful suggestions. This work has been supported by the MICINN of the Spanish government through the I3 Program and grant No. FIS2008-01299 co-financed by European Union's FEDER funds. We also thank the support of the "Spanish Interdisciplinary Network on the Biophysics of Viruses" (BioFiVinet) (grants No. FIS2010-10552-E and FIS2011-16090-E).

References and Further Reading

1. Crick F, Watson J (1956) Structure of small viruses. Nature 177:473–475
2. Caspar D, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24
3. Luque A, Reguera D (2010) The structure of elongated viral capsids. Biophys J 98:2993–3003
4. Lidmar J, Mirny L, Nelson DR (2003) Virus shapes and buckling transitions in spherical shells. Phys Rev E 68:051910
5. Twarock R (2006) Mathematical virology: a novel approach to the structure and assembly of viruses. Philos Trans A Math Phys Eng Sci 364:3357–3373
6. Zandi R, Reguera D, Bruinsma RF et al (2004) Origin of icosahedral symmetry in viruses. Proc Natl Acad Sci USA 101:15556–15560
7. Luque A, Zandi R, Reguera D (2010) Optimal architectures of elongated viruses. Proc Natl Acad Sci USA 107:5323–5328
8. Fejer S, James T, Hernandez-Rojas J, Wales D (2009) Energy landscapes for shells assembled from pentagonal and hexagonal pyramids. Phys Chem Chem Phys 11:2098–2104
9. Fejer S, Chakrabarti D, Wales D (2010) Emergent complexity from simple anisotropic building blocks: shells, tubes, and spirals. ACS Nano 4:219–228
10. Rapaport DC (2010) Modeling capsid self-assembly: design and analysis. Phys Biol 7:045001
11. Elrad OM, Hagan MF (2010) Encapsulation of a polymer by an icosahedral virus. Phys Biol 7:045003
12. Nguyen HD, Reddy VS, Brooks CL (2009) Invariant polymorphism in virus capsid assembly. J Am Chem Soc 131:2606–2614
13. Arkhipov A, Freddolino PL, Schulten K (2006) Stability and dynamics of virus capsids described by coarse-grained modeling. Structure 14:1767–1777
14. Freddolino PL, Arkhipov AS, Larson SB et al (2006) Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure 14:437–449
15. Keef T, Twarock R (2008) New insights into viral architecture *via* affine extended symmetry groups. Comput Math Meth Med 9:221–229
16. Bruinsma RF, Gelbart WM, Reguera D et al (2003) Viral self-assembly as a thermodynamic process. Phys Rev Lett 90:248101
17. Landau LD, Lifshitz EM (1975) Theory of Elasticity. Pergamon Press, London
18. Nguyen TT, Bruinsma RF, Gelbart WM (2005) Elasticity theory and shape transitions of viral shells. Phys Rev E 72:051923
19. Klug WS, Bruinsma RF, Michel J-P, Knobler CM (2006) Failure of viral shells. Phys Rev Lett 97:228101
20. Luque A (2011) Structure, mechanical properties, and self-assembly of viral capsids. Ph.D. Thesis, University of Barcelona
21. Chen T, Glotzer S (2007) Simulation studies of a phenomenological model for elongated virus capsid formation. Phys Rev E 75:051504
22. Saunders M, Voth G (2012) Coarse-graining of multiprotein assemblies. Curr Opin Struct Biol 20:144–150
23. Krishna V, Ayton GS, Voth GA (2010) Role of protein interactions in defining HIV-1 viral capsid shape and stability: a coarse-grained analysis. Biophys J 98:18–26
24. Chen B, Tycko R (2011) Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation. Biophys J 100:3035–3044
25. Miao Y, Johnson JE, Ortoleva PJ (2010) All-atom multiscale simulation of cowpea chlorotic mottle virus capsid swelling. J Phys Chem B 114:11181–11195
26. Devkota B, Petrov AS, Lemieux S et al (2009) Structural and electrostatic characterization of pariacoto virus: implications for viral assembly. Biopolymers 91:530–538
27. Lavelle L, Gingery M, Phillips M et al (2009) Phase diagram of self-assembled viral capsid protein polymorphs. J Phys Chem B 113:3813–3819
28. Hu Y, Zandi R, Anavitarte A et al (2008) Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size. Biophys J 94:1428–1436

29. Dixit SK, Goicochea NL, Daniel M-C et al (2006) Quantum dot encapsulation in viral capsids. Nano Lett 6:1993–1999
30. Kegel WK, van der Schoot P (2004) Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. Biophys J 86:3905–3913
31. Verwey EJW, Overbeek JTG (1999) Theory of stability of lyophobic colloids. Dover Press, Minneola
32. Siber A, Božič AL, Podgornik R (2012) Energies and pressures in viruses: contribution of nonspecific electrostatic interactions. Phys Chem Chem Phys 14:3746–3765
33. Ceres P, Zlotnick A (2002) Weak protein−protein interactions are sufficient to drive assembly of hepatitis B virus capsids. Biochemistry 41:11525–11531
34. Singh S, Zlotnick A (2003) Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. J Biol Chem 278:18249–18255
35. Zlotnick A (2003) How does your virus grow? Understanding and interfering with virus assembly. Trends Biotechnol 21:536–542
36. Casini GL, Graham D, Heine D et al (2004) *In vitro* papillomavirus capsid assembly analyzed by light scattering. Virology 325:320–327
37. Chen C, Kao CC, Dragnea B (2008) Self-assembly of brome mosaic virus capsids: insights from shorter time-scale experiments. J Phys Chem A 112:9405–9412
38. Zandi R, van der Schoot P, Reguera D et al (2006) Classical nucleation theory of virus capsids. Biophys J 90:1939–1948
39. Schwartz R, Shor PW, Prevelige PE, Berger B (1998) Local rules simulation of the kinetics of virus capsid self-assembly. Biophys J 75:2626–6636
40. Berger B, Shor P, Tucker-Kellogg L, King J (1994) Local rule-based theory of virus shell assembly. Proc Natl Acad Sci USA 91:7732–7736
41. Keef T, Micheletti C, Twarock R (2006) Master equation approach to the assembly of viral capsids. J Theor Biol 242:713–721
42. Moisant P, Neeman H, Zlotnick A (2010) Exploring the paths of (virus) assembly. Biophys J 99:1350–1357
43. Wilber AW, Doye JPK, Louis AA et al (2007) Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. J Chem Phys 127:085106
44. Hagan MF, Chandler D (2006) Dynamic pathways for viral capsid assembly. Biophys J 91:42–54
45. Hagan MF, Elrad OM (2010) Understanding the concentration dependence of viral capsid assembly kinetics–the origin of the lag time and identifying the critical nucleus size. Biophys J 98:1065–1074
46. Morozov A, Rudnick J, Bruinsma R, Klug W (2010) Assembly and disassembly of deltahedral viral shells. In: Stockley P, Twarock R (eds) Emerging topics in physical virology. Imperial College Press, London, pp 159–183
47. Levandovsky A, Zandi R (2009) Nonequilibirum assembly, retroviruses, and conical structures. Phys Rev Lett 102:198102
48. Hicks SD, Henley CL (2006) Irreversible growth model for virus capsid assembly. Phys Rev E 74:031912
49. Bruinsma RF (2006) Physics of RNA and viral assembly. Eur Phys J E Soft Matter 19:303–310
50. Ting CL, Wu J, Wang Z-G (2011) Thermodynamic basis for the genome to capsid charge relationship in viral encapsidation. Proc Natl Acad Sci USA 108:16986–16991
51. van der Schoot P, Bruinsma R (2005) Electrostatics and the assembly of an RNA virus. Phys Rev E 71:061928
52. Zandi R, van der Schoot P (2009) Size regulation of ss-RNA viruses. Biophys J 96:9–20
53. Cadena-Nava RD, Comas-Garcia M, Garmann RF et al (2012) Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J Virol 86:3318–3326
54. ElSawy KM, Caves LSD, Twarock R (2011) On the origin of order in the genome organization of ssRNA viruses. Biophys J 101:774–780
55. Mahalik JP, Muthukumar M (2012) Langevin dynamics simulation of polymer-assisted virus-like assembly. J Chem Phys 136:135101

56. Harvey S, Petrov A, Devkota B (2009) Viral assembly: a molecular modeling perspective. Phys Chem Chem Phys 11:10541–10542
57. Odijk T (1998) Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. Biophys J 75:1223–1227
58. Purohit PK, Inamdar MM, Grayson PD et al (2005) Forces during bacteriophage DNA packaging and ejection. Biophys J 88:851–866
59. Kindt J, Ben-Shaul A, Tzlil S, Gelbart WM (2001) DNA packaging and ejection forces in bacteriophage. Proc Natl Acad Sci USA 98:13671–13674
60. Rau D, Lee B, Parsegian V (1984) Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. Proc Natl Acad Sci USA 81:2621–2625
61. Inamdar MM, Gelbart WM, Phillips R (2006) Dynamics of DNA ejection from bacteriophage. Biophys J 91:411–420
62. Smith DE, Tans SJJ, Smith SBB et al (2001) The bacteriophage phi29 portal motor can package DNA against a large internal force. Nature 413:748–752
63. Evilevitch A, Lavelle L, Knobler CM et al (2003) Osmotic pressure inhibition of DNA ejection from phage. Proc Natl Acad Sci USA 100:9292–9295
64. Cordova A, Deserno M, Gelbart WM, Ben-Shaul A (2003) Osmotic shock and the strength of viral capsids. Biophys J 85:70–74
65. Roos WH, Bruinsma R, Wuite GJL (2010) Physical virology. Nat Phys 6:733–743
66. Vella D, Ajdari A, Vaziri A, Boudaoud A (2012) The indentation of pressurized elastic shells: from polymeric capsules to yeast cells. J R Soc Interface 9:448–455
67. Carrasco C, Luque A, Hernando-Pérez M et al (2011) Built-in mechanical stress in viral shells. Biophys J 100:1100–1108
68. Purohit PK, Kondev J, Phillips R (2003) Mechanics of DNA packaging in viruses. Proc Natl Acad Sci USA 100:3173–3178
69. Zandi R, Reguera D (2005) Mechanical properties of viral capsids. Phys Rev E 72:021917
70. Aznar M, Luque A, Reguera D (2012) Relevance of capsid structure in the buckling and maturation of spherical viruses. Phys Biol 9:036003
71. Buenemann M, Lenz P (2008) Elastic properties and mechanical stability of chiral and filled viral capsids. Phys Rev E 78:051924
72. Vliegenthart GA, Gompper G (2006) Mechanical deformation of spherical viruses with icosahedral symmetry. Biophys J 91:834–841
73. Roos WH, Gibbons MM, Arkhipov A et al (2010) Squeezing protein shells: how continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale. Biophys J 99:1175–1181
74. Arkhipov A, Roos WH, Wuite GJL, Schulten K (2009) Elucidating the mechanism behind irreversible deformation of viral capsids. Biophys J 97:2061–2069
75. Tama F, Wriggers W, Brooks CL (2002) Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J Mol Biol 321:297–305
76. Gibbons MM, Klug WS (2007) Mechanical modeling of viral capsids. J Mater Sci 42:8995–9004
77. May ER, Feng J, Brooks CL (2012) Exploring the symmetry and mechanism of virus capsid maturation *via* an ensemble of pathways. Biophys J 102:606–612
78. May E, Brooks C (2011) Determination of viral capsid elastic properties from equilibrium thermal fluctuations. Phys Rev Lett 106:188101
79. Roos W, Gertsman I, May E et al (2012) Mechanics of bacteriophage maturation. Proc Natl Acad Sci USA 109:2342–2347
80. Olson G (1999) New directions in martensite theory. Mater Sci Eng A-Struct 273–275:11–20
81. Falk W, James R (2006) Elasticity theory for self-assembled protein lattices with application to the martensitic phase transition in bacteriophage T4 tail sheath. Phys Rev E 73:011917
82. Zandi R, Reguera D, Rudnick J, Gelbart WM (2003) What drives the translocation of stiff chains? Proc Natl Acad Sci USA 100:8649–8653

83. Evilevitch A (2006) Effects of condensing agent and nuclease on the extent of ejection from phage lambda. J Phys Chem B 110:22261–22265
84. Grayson P, Han L, Winther T, Phillips R (2007) Real-time observations of single bacteriophage lambda DNA ejections *in vitro*. Proc Natl Acad Sci USA 104:14652–14657
85. Simons K, Garoff H (1980) The budding mechanisms of enveloped animal viruses. J Gen Virol 50:1–21
86. Lerner DM, Deutsch JM, Oster GF (1993) How does a virus bud? Biophys J 65:73–79
87. Deserno M (2004) Elastic deformation of a fluid membrane upon colloid binding. Phys Rev E 69:031903
88. Ruiz-Herrero T, Velasco E, Hagan MF (2012) Mechanisms of budding of nanoscale particles through lipid bilayers. J Phys Chem B 116:9595–9603
89. Zhang R, Nguyen TT (2008) Model of human immunodeficiency virus budding and self-assembly: role of the cell membrane. Phys Rev E 78:051903
90. Tzlil S, Deserno M, Gelbart WM, Ben-Shaul A (2004) A statistical-thermodynamic model of viral budding. Biophys J 86:2037–2048
91. Gao H, Shi W, Freund LB (2005) Mechanics of receptor-mediated endocytosis. Proc Natl Acad Sci USA 102:9469–9474

Further Reading

Israelachvili J (2011) Intermolecular and surface forces, 3rd edn. Academic Press, London
Katen S, Zlotnick A (2009) The thermodynamics of virus capsid assembly. Methods Enzymol 455:395–417
Knobler CM, Gelbart WM (2009) Physical chemistry of DNA viruses. Annu Rev Phys Chem 60:367–383
Marenduzzo D, Micheletti C, Orlandini E (2010) Biopolymer organization upon confinement. J Phys Condens Matter 22:283102
Nurmemmedov E, Castelnovo M, Catalano CE, Evilevitch A (2007) Biophysics of viral infectivity: matching genome length with capsid size. Q Rev Biophys 40:327–356
Phillips R, Kondev J, Theriot J, Orme N (2010) Physical biology of the cell. Garland Science, New York
Stockley PG, Twarock R (2010) Emerging topics in physical virology. Imperial College Press, London

Also especially recommended for further reading are references [18, 20, 30, 32, 35, 58, 65, 71] *listed above*.

Part IV
Applied Structural and Physical Virology

Chapter 20
Antiviral Agents: Structural Basis of Action and Rational Design

Luis Menéndez-Arias and Federico Gago

Abstract During the last 30 years, significant progress has been made in the development of novel antiviral drugs, mainly crystallizing in the establishment of potent antiretroviral therapies and the approval of drugs inhibiting hepatitis C virus replication. Although major targets of antiviral intervention involve intracellular processes required for the synthesis of viral proteins and nucleic acids, a number of inhibitors blocking virus assembly, budding, maturation, entry or uncoating act on virions or viral capsids. In this review, we focus on the drug discovery process while presenting the currently used methodologies to identify novel antiviral drugs by using a computer-based approach. We provide examples illustrating structure-based antiviral drug development, specifically neuraminidase inhibitors against influenza virus (*e.g.* oseltamivir and zanamivir) and human immunodeficiency virus type 1 protease inhibitors (*i.e.* the development of darunavir from early peptidomimetic compounds such as saquinavir). A number of drugs in preclinical development acting against picornaviruses, hepatitis B virus and human immunodeficiency virus and their mechanism of action are presented to show how viral capsids can be exploited as targets of antiviral therapy.

Keywords Antiretroviral drugs • Capsid proteins • DNA polymerases • Drug development • Fusion inhibitors • Hepatitis virus • Herpesviruses • Human

L. Menéndez-Arias (✉)
Centro de Biología Molecular "Severo Ochoa" (Consejo Superior de Investigaciones Científicas & Universidad Autónoma de Madrid), c/Nicolás Cabrera 1, Campus de Cantoblanco, 28049 Madrid, Spain
e-mail: lmenendez@cbm.uam.es

F. Gago (✉)
Department of Pharmacology, Universidad de Alcalá, Alcalá de Henares, 28871 Madrid, Spain
e-mail: federico.gago@uah.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_20,

immunodeficiency virus • Influenza virus • Ligand docking • Neuraminidase • Nucleoside analogues • Proteases • Viral assembly • Viral entry • Viral replication • Virtual screening

Abbreviations

AIDS	Acquired immune deficiency syndrome
CoMFA	Comparative molecular field analysis
CTD	C-terminal domain
HBV	Hepatitis B virus
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HR	Heptad repeat
HSV	Herpes simplex virus
HTS	High-throughput screening
LBVS	Ligand-based virtual screening
mRNA	Messenger RNA
Neu5Ac	*N*-acetylneuraminic acid
NMR	Nuclear magnetic resonance
NNRTIs	Nonnucleoside RT inhibitors
NTD	N-terminal domain
PDB	Protein Data Bank
RSV	Respiratory syncytial virus
RT	Reverse transcriptase
SBVS	Structure-based virtual screening
THF	Tetrahydrofuran
VZV	Varicella-zoster virus

20.1 Introduction

Antiviral drugs are compounds that stop the development and propagation of a virus without causing a relevant damage in the host cell. Despite landmark achievements (*e.g.* >30 new drugs approved during the last three decades to fight AIDS), the number of available antiviral compounds is still small and effective only against a limited group of pathogens. Examples are human immunodeficiency virus (HIV), herpes simplex virus (HSV), varicella-zoster virus (VZV), human cytomegalovirus (HCMV), influenza virus and hepatitis B and C viruses (HBV and HCV, respectively) [1].

There are several reasons that account for the difficulties in developing antiviral agents. First, viruses are obligatory intracellular parasites. Because every step in the

viral life cycle engages host functions, it is difficult to interfere with virus growth without having a negative impact on the host cell. Side effects are relatively common. Second, clinically important viruses are often dangerous and cannot be propagated or tested in model systems. Thus, viruses causing fatal diseases in humans (*e.g.* smallpox or several hemorrhagic diseases) have to be handled by well-trained and experienced scientists, and in facilities with strict containment requirements. Not surprisingly, these labs are expensive and difficult to maintain. Apart from the biological safety limitations, sometimes viral infections cannot be properly monitored for antiviral drug development due to the lack of appropriate animal models of human disease (*e.g.* smallpox or measles) or to difficulties in growing the virus in cell culture (*e.g.* HBV).

A third factor that limits the efficacy of antiviral drugs is their potency requirements. Ideally, an antiviral agent should be extremely potent. Partial inhibition is not acceptable for an antiviral drug. The reason is that limited viral replication under drug pressure allows for the generation of variants that can be selected under treatment. The emergence of resistance is a major drawback of many antiviral therapies. For example, in the case of HIV, therapies prescribed in the late 1980s or early 1990s were based on a single drug (mostly zidovudine) or combinations of two drugs (usually two inhibitors of the viral polymerase) [2]. However, those drugs were not potent enough to limit the emergence of drug-resistant variants [3] and therefore these viruses were almost impossible to combat successfully with the available drug armamentarium. In a patient with full blown AIDS, HIV production has been estimated at about 10^{12} virions per day. The high replication rate of HIV and the low fidelity of its DNA polymerase [4] trigger the appearance of drug resistance under suboptimal therapies.

Another issue that deserves some attention is the short duration of many viral infections (*e.g.* flu, common cold, etc. . . .). Very often, the symptoms appear when the virus is no longer replicating and are due to the immune response of the patient. In those cases, antiviral drugs should be administered early in infection or as a prophylactic measure in populations at risk. However, this could be potentially harmful for healthy populations.

20.2 Drug Discovery and Potential Targets of Antiviral Intervention

Early evidence of activity against vaccinia virus was reported for several thiosemicarbazone derivatives [5], and one of them (N-methylisatin-β-thiosemicarbazone) entered clinical studies for the prophylaxis of smallpox [6] at the time when vaccination against smallpox virus took over. The first antiviral drug licensed for clinical use was a thymidine analogue known as idoxuridine (5′-iodo-2′-deoxyuridine), whose synthesis was described in 1959 [7]. Idoxuridine has been used topically to treat eye and skin infections caused by herpes simplex virus. The drug acts on viral

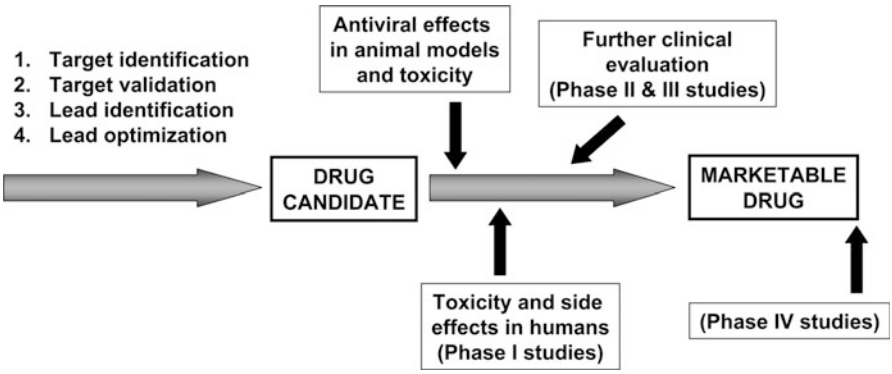


Fig. 20.1 Schematic outline of the drug discovery process

replication by interfering with the normal function of thymidylate phosphorylase and viral DNA polymerases.

Encouraged by the success with antibiotics in the 60s and 70s, drug companies launched huge blind-screening programs that were not very fruitful. Today, recombinant DNA technology and sophisticated chemistry [8], as well as impressive advances in structural and functional genomics, have facilitated the identification and analysis of particular proteins or mechanisms. Essential viral genes can be cloned and expressed in appropriate organisms so that the encoded proteins can be purified and analyzed in molecular and atomic detail.

Drug discovery programs start with the identification of suitable drug targets (Fig. 20.1). A drug target can be defined as a biomolecule (usually a single protein or a protein complex) linked to a disease and containing a suitable binding site that can be exploited to modulate its function. These targets need to be validated to demonstrate that they are critically involved in a disease and that their modulation is likely to have a therapeutic effect. In virology, many drug targets are viral proteins (*e.g.* enzymes), nucleic acids or other biological macromolecules required in the virus life cycle. Infection and viral propagation can be blocked by small compounds binding to relevant targets. Once the target has been validated *in vitro* and/or in animal models, lead identification starts with the design and development of a suitable assay to monitor the biological function under study. Active compounds that demonstrate dose-dependent target modulation and some degree of selectivity for the target under study are called lead compounds. These molecules are optimized in terms of potency and selectivity to become drug candidates.

In order to become a marketable drug, the candidate undergoes additional preclinical evaluation, including pharmacokinetic and toxicity studies in animal models. Clinical trials involving new drugs are commonly classified into four phases. Approval is usually granted if the drug advances successfully through phases I, II and III. The main objective of phase I is to assess drug safety and pharmacokinetics in a relatively small number of healthy individuals who receive small doses of the compound. In phase II, the testing protocol is established. These are trials trying to find out

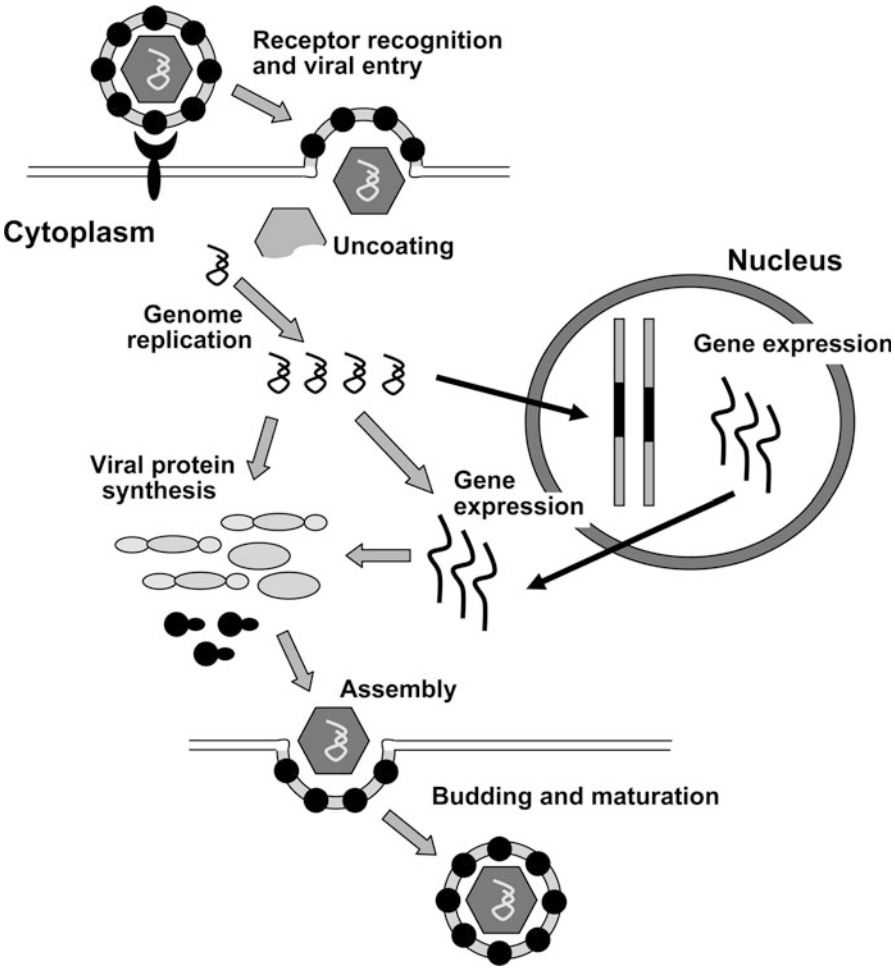


Fig. 20.2 Basic steps of viral replication as potential targets of antiviral therapy

appropriate doses, prescription regimens, etc.... and are carried out with patients. Phase III constitutes the final testing before approval and these trials try to show whether the drug has a measurable benefit or advantage over other treatments with a relatively large number of patients (usually >100 patients for antiviral drugs). Phase IV includes all studies carried out after approval of the drug and may address many different issues, from side-effects to efficacy in comparison with other drugs or regimens, and usually involve a large number of patients worldwide.

Recent years have witnessed a remarkable progress in the discovery and development of antiviral agents, fueled by advances in the understanding of viral life cycles which in turn have provided new opportunities for therapeutic intervention. Although the mechanisms of replication and propagation can show significant variations between different viruses, a prototypic life cycle is presented in

Fig. 20.2 (see also Chap. 1). Typically, a virion (*i.e.* an infectious viral particle) first attaches to the surface of the host cell. This interaction can be specific and involve the participation of one or more different types of proteins. For example, in HIV-1, the main receptor is CD4, but other proteins (*e.g.* chemokine receptors CCR5, CXCR4, etc. . . .) can facilitate viral entry acting as coreceptors (see Chap. 15).

Viral uptake occurs through different mechanisms such as endocytosis (as for example, in the case of influenza virus) or the fusion of the cellular membrane with the viral envelope, as demonstrated in the case of retroviruses (see Chap. 16). These events allow the internalization of the capsid containing the viral genome. Uncoating or disassembly is a still poorly understood process that releases the viral genome from its protein shell (see Chaps. 15 and 16). Disassembly may occur rapidly after fusion, as occurs in HIV, or be triggered by pH variations, as observed in viruses entering the cell by an endocytic pathway (*e.g.* influenza virus). The replication of the viral genome occurs by different mechanisms depending on whether the viral genome is a single- or double-stranded DNA or RNA. In general, DNA viruses use cellular pathways to replicate their genomes, but RNA viruses (and in general, those replicating in the cytoplasm) provide their own enzymes to complete virus replication. Cellular factors, together with specific proteins encoded within the viral genome, contribute to transcription and post-transcriptional modification of viral messenger RNAs (mRNAs). In some viruses (for example, in retroviruses), viral mRNAs are generated in the nucleus by the action of the cellular RNA polymerase. This is due to the integration of the viral genome (as a double-stranded DNA) into the genome of the host cell.

Viral polyproteins are synthesized by the cellular translational machinery. After post-translational modification, viral proteins and their genomes are transported to assembly sites within the host cell (see Chap. 14). Viral factors may also participate in these processes. The sites of assembly are frequently located at the plasma membrane or in intracellular factories (often associated with membranes). Virion assembly is a complex process which involves multiple molecular recognition steps and conformational transitions. The viral capsid is assembled in a multimerization reaction, with or without the help of scaffolding proteins or viral nucleic acids (see Chaps. 10, 11 and 19). The viral nucleic acids are packaged into the capsid during or after its assembly (see Chap. 12). Finally, the viral particle can be transformed into an infectious virion through a maturation process that involves changes in the structure and properties of the capsid (see Chap. 13).

In enveloped viruses, a membrane with embedded viral proteins is incorporated into the virion (see Chap. 11), and the resulting viral particles are released after budding. In contrast, cell lysis mediates the release of non-enveloped viruses. In some cases (*e.g.* in retroviruses), maturation occurs once the immature virion has been released and involves the proteolytic processing of precursors containing the viral proteins, including those that form the capsid.

All of those steps of the viral life cycle constitute potential targets of antiviral intervention. At present, drugs inhibiting viral enzymes involved in the replication or expression of the viral genome are commonly used in antiviral treatments. However, recent developments including the determination of structures, properties and functions of capsids and virions, as well as the elucidation of events involving

interactions between components of the viral particle or between them and host cell molecules (the subjects of this book) have opened novel avenues for the design of drugs acting directly on the viral particle. This chapter provides examples of approved or pre-clinical antiviral strategies directed at inhibiting viral nucleic acid metabolism, as well as others aimed at interfering with cell recognition, entry, uncoating, assembly, or maturation of virus particles.

20.3 Antiviral Drugs and Mechanisms of Action

Licensed compounds used in the treatment of viral infections target HIV, HBV, HCV, influenza virus, HSV, and other herpesviruses such as VZV and HCMV. A number of drugs act on steps that lead to the formation of the viral capsid or the mature virion (*i.e.* assembly, budding and maturation) while others, whose target is the assembled capsid or the virion, interfere with processes affecting viral entry and uncoating [1]. Nonetheless, most of the approved drugs block intracellular events affecting the synthesis and dynamics of viral proteins and nucleic acids (Table 20.1). Within this group, viral polymerases constitute the major target for many antiviral drugs.

20.3.1 Drugs Blocking Intracellular Processes Required for the Synthesis of Viral Components

Viral Genome Replication Inhibitors

Compounds inhibiting the replication of HSV, VZV and HCMV include prodrugs of nucleoside analogues (*e.g.* valacyclovir, valganciclovir and famciclovir) that need to be phosphorylated in order to become substrates of the viral DNA polymerase (Fig. 20.3). Viral enzymes (*e.g.* thymidine kinases in HSV and VZV, and a protein kinase in HCMV) are responsible for the transformation of acyclovir, ganciclovir and penciclovir into their monophosphate derivatives. Further phosphorylation steps are carried out by host cell kinases. The triphosphate derivatives of acyclovir, ganciclovir and penciclovir that mimic the natural substrates of the viral DNA polymerase, are incorporated into the growing DNA chain and often terminate viral replication. because they lack a 3′-OH in their ribose ring. Cidofovir is a phosphonate-containing acyclic cytosine analogue that, unlike the inhibitors described above, does not depend on viral enzymes for its conversion to the triphosphorylated form that competes with the dNTP substrates [9]. Foscarnet is an analogue of pyrophosphate, a product of the nucleotide incorporation reaction, and therefore behaves as an inhibitor of DNA polymerization. Unfortunately, our understanding of the mechanisms involved in resistance to acyclovir and other related inhibitors is limited by the absence of crystal structures of herpesvirus DNA polymerases.

Table 20.1 Antiviral drugs approved for clinical use

Target	Step of the viral life cycle or cellular function inhibited	Virus	Drug type and name
Viral	Entry	HIV	Fusion inhibitors: Enfuvirtide
	Disassembly/uncoating	Influenza virus	Drugs binding to the viral protein M2 (an ion channel): Amantadine and rimantadine
	Genome replication	HIV	Nucleoside/nucleotide RT inhibitors: Zidovudine (AZT), didanosine (ddI), zalcitabine (ddC), stavudine (d4T), lamivudine (3TC), abacavir (ABC), emtricitabine (FTC) and tenofovir (tenofovir disoproxil fumarate)[a]
			Nonnucleoside RT inhibitors: Nevirapine, delavirdine, efavirenz, etravirine and rilpivirine
		HBV	Nucleoside/nucleotide analogues: Lamivudine, emtricitabine, entecavir, telbivudine, adefovir (adefovir dipivoxil)[a] and tenofovir (tenofovir disoproxil fumarate)[a]
		HSV and VZV	Nucleoside/nucleotide analogues: Acyclovir (valaciclovir),[a] penciclovir (famciclovir),[a] idoxuridine, trifluridine and brivudine
		HCMV	Nucleoside/nucleotide analogues: Ganciclovir (valganciclovir) and cidofovir
			Pyrophosphate analogue: Foscarnet
	Integration into the host genome	HIV	HIV integrase inhibitors: Raltegravir and elvitegravir
	Synthesis of viral mRNAs	HCMV	Antisense (phosphorothioate) oligonucleotide: Fomivirsen
	Cleavage of viral polyproteins	HIV	HIV protease inhibitors: Saquinavir, ritonavir, indinavir, nelfinavir, amprenavir and its prodrug fosamprenavir, lopinavir, atazanavir, tipranavir and darunavir
		HCV	HCV protease inhibitors: Telaprevir and boceprevir
	Budding	Influenza virus	Viral neuraminidase inhibitors: Oseltamivir and zanamivir
Cellular	Viral entry	HIV	Viral coreceptor inhibitors: Maraviroc
	Innate immunity	HCV and HBV	Interferons: Pegylated interferons α-2a and α-2b, and interferons α-2a and α-2b
	mRNA capping enzymes and viral mutagenesis	HCV and influenza virus	Ribonucleoside analogue: Ribavirin

[a]Compound approved as a pro-drug, whose name is indicated between parentheses

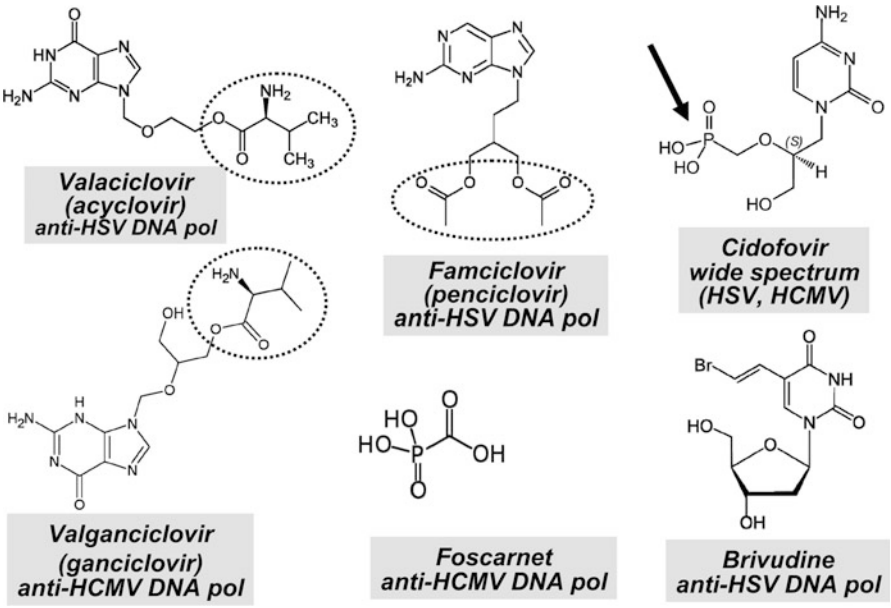


Fig. 20.3 Approved nucleoside, nucleotide and pyrophosphate analogues with antiviral effect on herpesviruses. Valaciclovir, famciclovir and valganciclovir are prodrugs of acyclovir, penciclovir and ganciclovir, respectively. Protecting groups that favor oral absorption are shown within *dotted ellipses*. *Arrow* shows the location of the phosphate group in the nucleotide analogue. *HSV* herpes simplex virus, *HCMV* human cytomegalovirus

HIV is a retrovirus that replicates through a proviral double-stranded DNA intermediate. Its polymerase, known as reverse transcriptase (RT), is able to synthesize DNA by using either RNA or DNA as templates. Nucleoside/nucleotide analogues and nonnucleoside RT inhibitors (NNRTIs) constitute the backbone of current antiretroviral therapies [2, 3]. Nucleoside/nucleotide RT inhibitors are prodrugs that need to be converted to active triphosphate analogues in order to be incorporated into the DNA during reverse transcription. They act as chain terminators thereby blocking DNA synthesis [10]. The HIV-1 RT is a heterodimer composed of subunits of 560 and 440 amino acids (known as p66 and p51, respectively). Both polypeptide chains have the same amino acid sequence, but the large subunit contains an RNase H domain that includes the 120 amino acids of its C-terminal region. The DNA polymerase domain is formed by residues 1–315 and aspartic acid residues 110, 185 and 186 in p66 constitute the catalytic triad.

There are many crystal structures available for HIV-1 RT, including relevant complexes such as the ternary complex of HIV-1 RT/double-stranded DNA/dTTP, or RT bound to RNA/DNA or to DNA/DNA template-primers. Additional structures containing DNA primers blocked with nucleoside analogues have also been described. A large number of NNRTIs have been crystallized in complex with the viral polymerase (for a review, see [11]). NNRTIs bind to a hydrophobic pocket

located about 8–10 Å away from the DNA polymerase active site. The first NNRTIs were identified by using an antiviral screening approach, but structure-based drug design has played a prominent role in the design and development of next-generation NNRTIs such as etravirine or rilpivirine [12, 13].

Reverse transcription is also involved in the replication of HBV. These viruses form an immature RNA-containing nucleocapsid that, inside the cell, is converted into a mature nucleocapsid containing a relaxed circular DNA [14]. The DNA found in extracellular virions is synthesized by viral RT using the pregenomic RNA as a template. The HBV RT shows sequence similarity with HIV-1 RT, and several inhibitors of HIV-1 replication (*e.g.* lamivudine, adefovir or tenofovir) have been approved from treating HBV infections (Table 20.1). Unfortunately, there is no available crystal structure for the HBV RT. The existence of this gap of knowledge can be attributed to major difficulties in obtaining a catalytically active polymerase.

RNA polymerases play a key role in the replication of important pathogenic RNA viruses such as HCV, poliovirus, rhinovirus, influenza virus and others. However, drugs targeting viral RNA polymerases have not received approval. At present, HCV RNA polymerase inhibitors are probably the ones in a more advanced stage of development [15, 16]. The HCV replicase (NS5B, an RNA-dependent RNA polymerase) is a 65-kDa protein that contains a C-terminal membrane insertion sequence that traverses the phospholipid bilayer as a transmembrane segment. NS5B inhibitors can be classified into nucleoside and nonnucleoside analogues. Examples of the first group are valopicitabine, 2′-*C*-methyladenosine, 4′-azidocytidine and mericitabine. The nucleotide prodrug β-D-2′-deoxy-2′-α-fluoro-2′-β-C-methyluridine (PSI-7977) [17] is currently in phase III clinical trials and showed promising results when combined with pegylated interferon and ribavirin. Nonnucleoside inhibitors of NS5B are tegobuvir, filibuvir and pyranoindoles such as HCV-371 and HCV-570 [18].

The non-structural 5A (NS5A) protein of HCV has been identified as the target of daclatasvir (BMS-790052), a thiazolidinone-containing small molecule that was shown to inhibit HCV replication in a cell-based replicon screen [19]. The inhibitor showed picomolar potency in preclinical assays. The enzymatic function of NS5A is not known, although it interacts with NS5B and modulates the host cell interferon response. Resistance to daclatasvir is associated with mutations in the N-terminal region of NS5A [20]. Interestingly, the cyclic endecapeptide alisporivir, currently in Phase III clinical trials, is a cyclophilin A-binding drug that selects for mutations in NS5A, although in a different domain [21].

Drugs Interacting with Other Targets

Gene expression is another possible target for antiviral intervention. A phosphorothioate derivative of the oligonucleotide 5′-GCG TTT GCT CTT CTT CTT GCG-3′ (known as fomivirsen) has been used as a repressor of the synthesis of early genes in HCMV infections. It binds to the complementary sequence in the viral mRNA, blocking its translation. In retroviruses, integration of the double-stranded proviral DNA (the product of reverse transcription) into the host cell DNA is essential for expression of viral mRNA. Retroviral mRNA is generated by transcription of the integrated provirus by cellular RNA polymerases.

Raltegravir and elvitegravir are inhibitors of HIV-1 integrase [22, 23], an enzyme endowed with 3′ endonucleolytic and strand transfer activities. It is composed of three structural domains: an N-terminal domain of unknown function, the central catalytic domain, and a C-terminal domain with DNA binding activity. Approved drugs bind to the central domain, close to the catalytic residues Asp64, Asp116 and Glu152, and act by blocking the strand transfer activity of the integrase. Crystal structures of catalytic domains have been used in the design of integrase inhibitors, but the full integrase (having all three domains) has become available only very recently [24].

Viral proteins are frequently synthesized as precursor polypeptides that are cleaved by viral proteases. Sometimes, this cleavage occurs before the assembly of the viral capsid, which is formed by the association of mature proteins into capsomers. For example, in poliovirus, rhinovirus and other picornaviruses, 60 protomers made of processed capsid proteins VP0, VP1 and VP3 assemble to form the icosahedral particle. In HCV, its serine protease (designated as NS3) interacts with an NS4A peptidic co-factor that modulates its proteolytic activity. Promising NS3/4A inhibitors were developed after rational structure-based design using a hexapeptide (Asp-Asp-Ile-Val-Pro-Cys) derived from the natural NS5A/5B cleavage site as a lead compound. After systematic replacement of single amino acid residues, important substituents in the protease binding pocket were identified and docked in the crystal structure of the unliganded NS3 protease. Further optimization led to the identification of ciluprevir (BILN-2061), one of the first promising drug candidates (Fig. 20.4). The recently licensed inhibitors boceprevir [25] and telaprevir [26] were developed by using similar approaches.

20.3.2 Inhibitors of Viral Assembly

Despite considerable efforts in recent years, drugs targeting the assembly of viral capsids (see Chaps. 10 and 11) have not yet gained approval for clinical use. At preclinical stages of development, there are interesting examples of drugs inhibiting the assembly of capsids of HBV (*e.g.* heteroaryldihydropyrimidines) [27, 28] and HIV-1 (CAI, CAP-1, bevirimat, etc. . . .) [29, 30] (see Sects. 20.6.2 and 20.6.3).

20.3.3 Budding Inhibitors

At present, antiviral treatment and prophylaxis of human influenza infections, including aggressive strains such as avian H_5N_1, are based on the use of oseltamivir and zanamivir that inhibit the release of infectious virus. The influenza virus membrane contains a glycoprotein (hemagglutinin) that recognizes sialic acid in cell-surface glycoproteins and glycolipids [31]. Sialic acid is the receptor for virus entry. However, newly formed virions may remain attached to the cell membrane

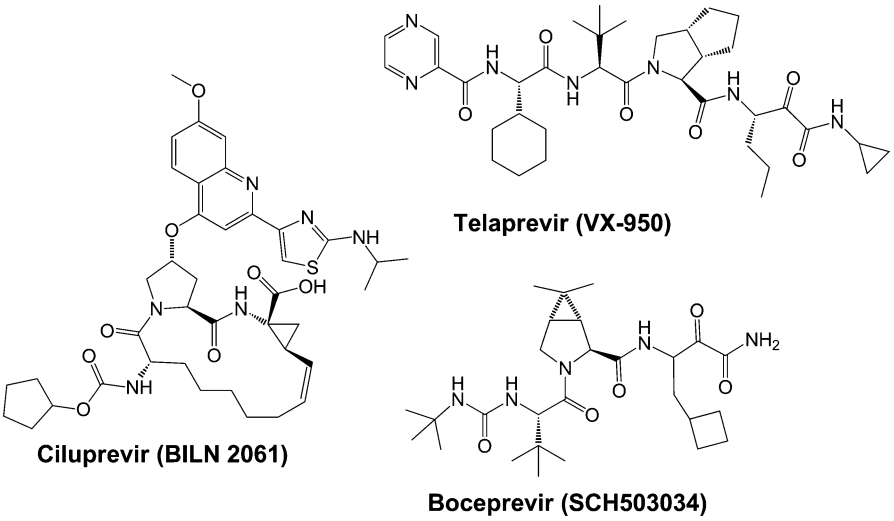


Fig. 20.4 Hepatitis C virus protease inhibitors

because of the interaction between hemagglutinin and sialic acid. The viral neuraminidase (also known as sialidase) facilitates the release of virions from infected cells by cleaving sialic acid residues. Oseltamivir and zanamivir inhibit this hydrolytic activity.

20.3.4 Maturation Inhibitors

Maturation is a relatively common process in animal and human viruses (see Chap. 13). In contrast to HCV and other virus families, specific cleavage of the polypeptides encoding the mature viral proteins occurs only on assembled immature particles. HIV-1 protease inhibitors block the proteolytic processing of precursor polypeptides Gag and Gag-Pol. Cleavage of Gag and Gag-Pol leads to the formation of structural proteins (*e.g.* MA, CA and NC), p6 and the viral enzymes protease, RT and integrase (Fig. 20.5). This process occurs extracellularly and is required for the formation of an infectious virion. The HIV-1 protease is a homodimeric enzyme composed of subunits of 99 amino acids with a symmetric substrate binding pocket. Inhibitors were initially designed as peptide derivatives mimicking the 'transition-state' and contained a non-hydrolysable bond at the position where the cleavage was expected to occur. All of the approved HIV protease inhibitors, except tipranavir, can be considered as peptidomimetics [2].

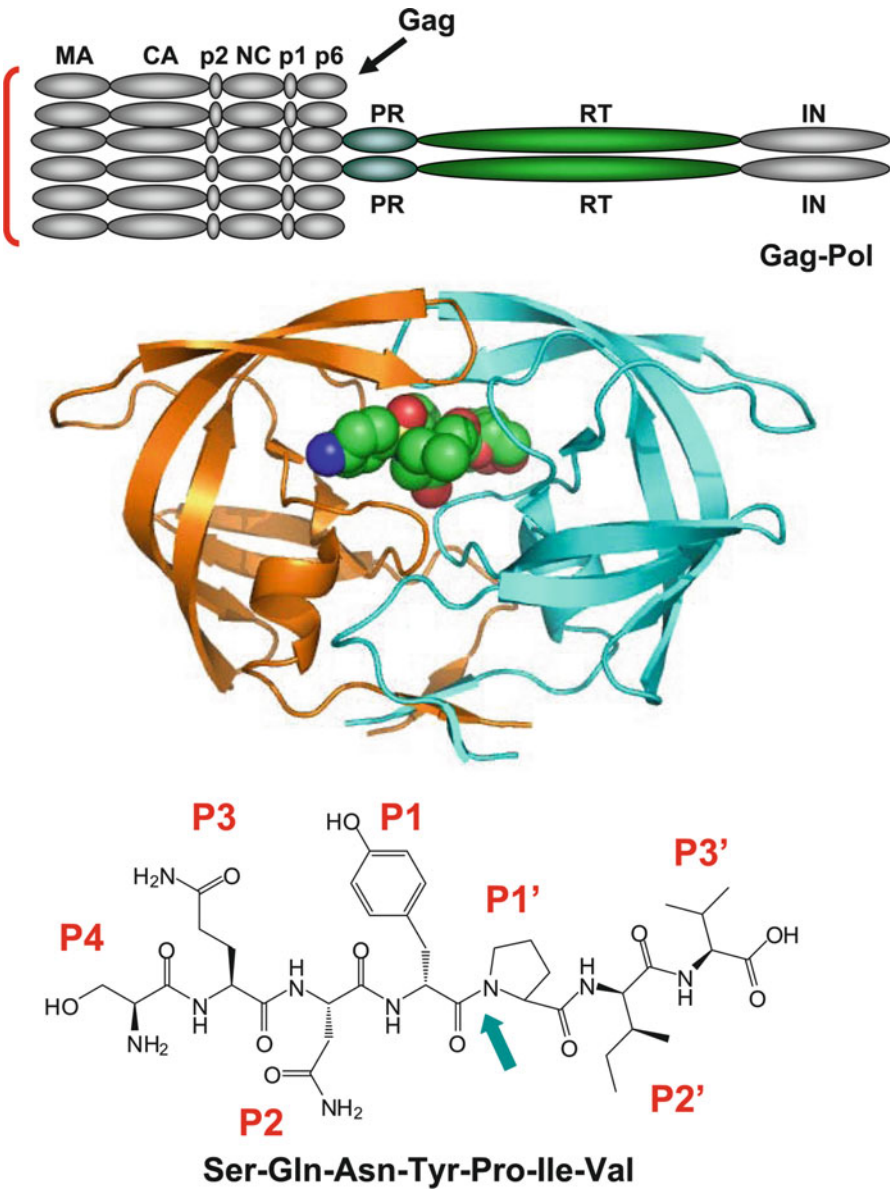


Fig. 20.5 Protein arrangement in Gag and Gag-Pol precursor polyproteins, HIV-1 protease structure and model substrate. Cartoon representation of the protease bound to darunavir (PDB entry 1T3R). Inhibitor C, O and N atoms are displayed as *green*, *red* and *purple spheres*, respectively. The minimal peptide substrate SQNY*PIV (asterisk indicates cleavage site) is shown below the protease structure

20.3.5 Drugs Targeting Viral Entry

Drugs acting at this step include compounds that block cell attachment, receptor recognition or the mechanisms leading to the translocation of the viral capsid to the host cell cytoplasm (see Chaps. 15 and 16). Two anti-HIV drugs belonging to this category have been approved: maraviroc and enfuvirtide. The first interferes with coreceptor recognition, while the second blocks the fusion of the viral envelope with the cell membrane.

Receptor and Coreceptor Recognition in HIV-1 Infection

HIV-1 enters target cells through a multi-step process that requires non-specific interactions with surface heparan sulfates, followed by binding of the viral envelope glycoprotein gp120 to its primary cellular receptor CD4. Attachment inhibitors in HIV-1 include dextran sulphate and other polyanions, as well as proteins binding glycosidic residues (*e.g.* mannose-specific lectins or cyanovirin N).

CD4 binds into a carbohydrate-free depression in gp120 that is poorly accessible to immunoglobulins. However, crystallographic studies carried out with complexes of gp120, CD4 and Fab fragments of neutralizing antibodies have shown that Phe43 of CD4 and residues 365–371 and 425–430 of gp120 make the largest contributions to interatomic contacts [32]. Phe43 is located on the CDR2-like loop of CD4 and binds within a hydrophobic pocket of gp120, termed the "Phe43 cavity", while Arg59 (in CD4) is located on a neighboring β-strand and forms an ion pair with the gp120 residue Asp368. The disruption of this interaction by small molecule binding has been validated as a promising approach to prevent HIV-1 entry. Inhibitors containing a tetramethylpiperidine ring such as NBD-556 displayed remarkable antiviral activity and altered the gp120 conformation in a similar way as observed upon CD4 binding [33, 34].

CCR5 and CXCR4 are the necessary coreceptors for cellular entry of HIV-1. These molecules are members of the seven-transmembrane G protein-coupled receptor superfamily, and can be blocked by natural ligands such as MIP-1α, MIP-1β and RANTES for CCR5 and SDF-1 for CXCR4. RANTES derivatives and SDF-1α have been investigated as potential inhibitors of viral entry. However, smaller CCR5 antagonists such as TAK-652 and TAK-779, aplaviroc, vicriviroc and maraviroc were further studied as potential anti-HIV agents. So far, maraviroc remains as the only clinically approved coreceptor antagonist. However, resistance to this drug appears mostly as a result of HIV-1 using CXCR4 as an alternate coreceptor for entry [3].

The most advanced CXCR4 antagonist, AMD3100 (Mozobil) has not been further studied as a therapeutic agent against HIV-1 [18]. However, in combination with granulocyte colony stimulating factor (and under the name of plerixafor), it has been approved for the mobilization of hematopoietic stem cells to the peripheral blood for collection and subsequent autologous transplantation in patients with non-Hodgkin's lymphoma and multiple myeloma.

Fusion Inhibitors: Targeting the Formation of Coiled-Coil Structures

Viral recognition by cellular receptors triggers a mechanism that leads to the fusion of the viral envelope and the target cell membrane [35] (see Chap. 16). Crystallographic studies of pre- and post-fusion conformations of the influenza virus hemagglutinin showed that the formation of a six-helix bundle played a major role in the fusogenic events. Drugs targeting the helix bundle have been developed to block entry of HIV-1 and respiratory syncytial virus (RSV).

In the case of HIV-1, enfuvirtide is a 36-amino acid peptide that derives from residues 127–162 of the transmembrane protein gp41. This glycoprotein has a trimeric structure, and each monomer contains a transmembrane region of 21 residues, located between an N-terminal ectodomain (175 residues) and a long cytoplasmic domain of 150 amino acids. The sequence of enfuvirtide overlaps with the heptad repeat two (HR2) region (residues 117–154) that interacts with the HR1 region (residues 29–82) of gp41. Enfuvirtide binds HR1 and prevents the formation of the trimeric coiled-coil structure required for membrane fusion. Despite showing efficacy *in vivo*, the clinical use of the drug has been limited by its reduced bioavailability and its large production costs. Novel anti-HIV drugs based on similar principles include tifuvirtide, sifuvirtide and T-2635. These peptides mimic in part the HR2 structure and exert their effects through a mechanism of action similar to that shown by enfuvirtide [35].

In the case of the human RSV, an orally active candidate (BMS-433771) was proposed to bind in a hydrophobic cavity within the trimeric N-terminal heptad repeat (equivalent to HR1), which is the region that associates with a C-terminal HR2 to form a six-helical coiled-coil bundle during the fusion process. X-ray crystallography of another drug candidate (TMC353121) bound to the human RSV fusion protein (Protein Data Bank (PDB) entry 3KPE) showed that this drug, rather than preventing formation of the coiled-coil, produces a local disturbance of the six-helix bundle by stabilizing the HR1 – HR2 interactions in an alternate conformation of the six-helix bundle [36].

20.3.6 Uncoating Inhibitors

Amantadine and rimantadine were used more than 20 years ago to prevent influenza A virus infections. However, they never gained wide acceptance and today virtually all strains are resistant to these agents. These compounds block the release of the viral capsid to the cytoplasm of the infected cell, because they interact with M2, an integral protein located in the viral envelope. It consists of four identical units, each one having three distinct domains: (i) an ectodomain made up of 24 residues exposed to the outside environment, (ii) a transmembrane region (mostly hydrophobic) composed of 22 amino acids, and (iii) 52 amino acids at the C-terminal region exposed to the intraviral milieu. M2 is activated by low pH and functions as an ion channel that facilitates entry of protons into the viral particle. Amantadine and rimantadine bind to the transmembrane region (Fig. 20.6), sterically blocking the channel and inhibiting the release of the viral capsid to the host cell cytoplasm.

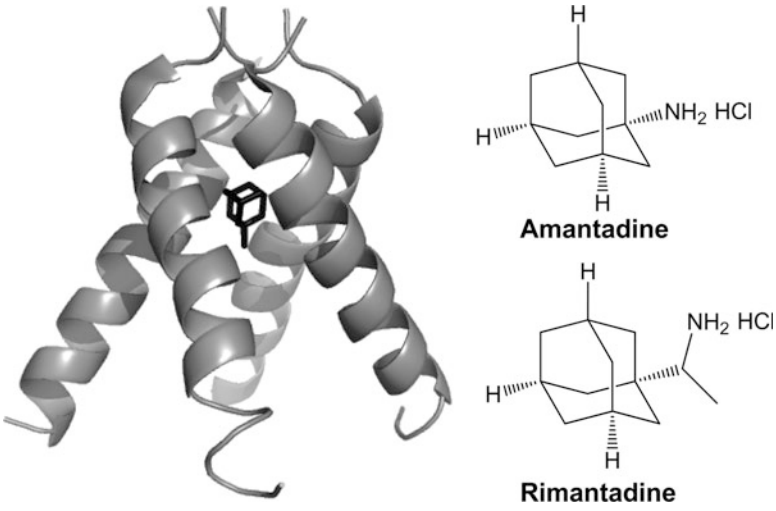


Fig. 20.6 Transmembrane region of influenza virus M2 protein bound to amantadine and chemical structures of drugs blocking the release of the viral capsid into the cytoplasm. Cartoon representation of the transmembrane regions of the M2 tetramer and stick model of the bound inhibitor (coordinates were taken from PDB entry 3C9J)

20.4 Strategies in the Development of Antiviral Drugs: From Random Screening to Structure-Based Design

Most of the antiviral drugs clinically used today and mentioned in previous sections of this chapter were discovered serendipitously or as a result of screening campaigns that made use of large libraries of synthetic compounds or natural products. In the traditional drug discovery process, it is estimated that screening of around 80,000 molecules can render one hit that needs to be further optimized to a lead compound by improving its activity and reducing its toxicity before entering clinical trials. Usually, these huge efforts are needed because of the lack of structures for ligands and protein targets. In this scenario, pharmaceutical companies invested in the development of high-throughput screening (HTS) assays and in the generation of large chemical libraries obtained by combinatorial chemistry. Nevertheless, this process is expensive, time-consuming and requires further efforts in target validation.

Random approaches including HTS and combinatorial chemistry are used in the absence of structural information about ligands or target proteins or macromolecular complexes. Over the last two decades we have been able to witness many profound changes in the drug discovery paradigm due to major advances in such diverse areas as HTS methods, genome sequence projects, macromolecular structure determination, quality and scope of publicly accessible databases, computer performance and new algorithms for *'in silico'* approaches [37] (Fig. 20.7). Among the latter, structure-based and ligand-based virtual screening (SBVS and LBVS, respectively) stand out as powerful tools that are being increasingly used in both

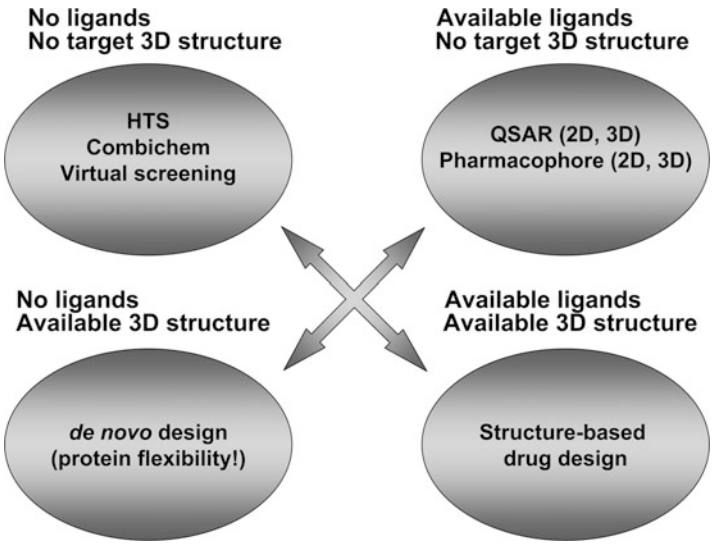


Fig. 20.7 Strategies in drug design depending on the availability of structural information for ligands and biological targets. Abbreviations: *2D* bidimensional, *3D* three-dimensional, *HTS* high-throughput screening, *QSAR* quantitative structure-activity relationship

industry and academia [38]. These techniques rely on the three-dimensional coordinates of either the macromolecular target (SBVS) or one or more known ligands (LBVS) to search in chemical or fragment libraries for putative hits that can then be tested experimentally for confirmation of affinity/activity and then eventually transformed into lead compounds. Due to the huge amount of ligands that can be found in currently available databases (*e.g.* 13 million in ZINC [39]), it is customary to employ a series of computational cost-effective filters to narrow down the number of molecules that will be subjected to virtual screening. Lipinski's rule of five[1] [40] and/or other physico-chemical-based principles can be used for filtering, as well as pharmacophoric hypotheses. One goal of LBVS is to find similarly shaped molecules that belong to novel chemotypes ("scaffold hopping") [41].

20.4.1 Structure-Based Drug Design

Structure-based drug design relies on knowledge of the three-dimensional structure of the biological target gained by means of techniques such as X-ray

[1] Lipinski's rule states that, in general, an orally active drug should meet at least three of the following criteria: (i) Not more than five hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms), (ii) not more than ten hydrogen bond acceptors (nitrogen or oxygen atoms), (iii) a molecular mass <500 daltons, and (iv) an octanol-water partition coefficient log P not greater than five.

crystallography (Chap. 4), nuclear magnetic resonance (NMR) spectroscopy (Chap. 5) and electron microscopy (Chap. 3). A number of computational methods have been developed to identify putative ligands that can fit into an active site or functional epitopes in the target protein or molecular complex. Success in molecular docking, which may be defined as an optimization problem that describes the "best-fit" orientation of a ligand within a binding site [42], very often depends on the quality of the three-dimensional target protein structure. Structures determined by X-ray crystallography at atomic or near-atomic resolution are the most adequate for docking. More specifically, successful identification of lead compounds usually requires a crystallographic resolution below 2.5 Å and an R factor well below 25 % (see Chap. 4).

20.4.2 Ligand Docking and Virtual Screening

Docking methods are calibrated and tested for their ability to predict, at the top of the list of possible solutions, the native conformation of a ligand bound to a protein as found in a representative series of crystallographic complexes. To assess the performance of a docking program in SBVS its ability to discriminate true binders from a pool of fake ligands ("decoys") is also examined. The goodness of fit between the ligand and the receptor is evaluated by means of a scoring function composed of different terms that attempt to account for the forces driving the binding event. Although the underlying physical laws describing the association process are well understood, accuracy and computational resources (mainly execution time) evolve in opposite directions, and fine tuning the appropriate balance between them is not an easy task. Therefore, accuracy is normally sacrificed for speed, especially in SBVS where very often too simplistic scoring functions are employed which plainly cannot capture the true free energy of binding that is measured experimentally [43]. Ligand conformers can be generated within the binding site itself "on the fly" for each target, or created beforehand as a collection of all major theoretically possible conformers following some pre-defined rules. The former method is more computer-intensive but, as an advantage, it can generate strained conformations that adapt better to each specific binding site environment. The latter method allows the stored coordinates to be reused again as many times as needed in different SBVS campaigns but the drawback is that a relevant conformation that deviates slightly from a canonical one can be missed.

Fragments [molecular weight $\leq$ 250 Da, log P $\leq$ 2.5, (where P refers to the octanol-water partition coefficient) and less than 6 rotatable bonds] are usually endowed with reduced affinities compared to conventional ligands but are better suited for chemical modifications aimed at producing novel drug candidates [44]. They can sample chemical space more effectively than do regular ligands. It is generally accepted that fragment docking outperforms traditional small-molecule docking in terms of hit rates even though erroneous binding modes can be obtained for them because their volumes are much smaller than those of typical binding site

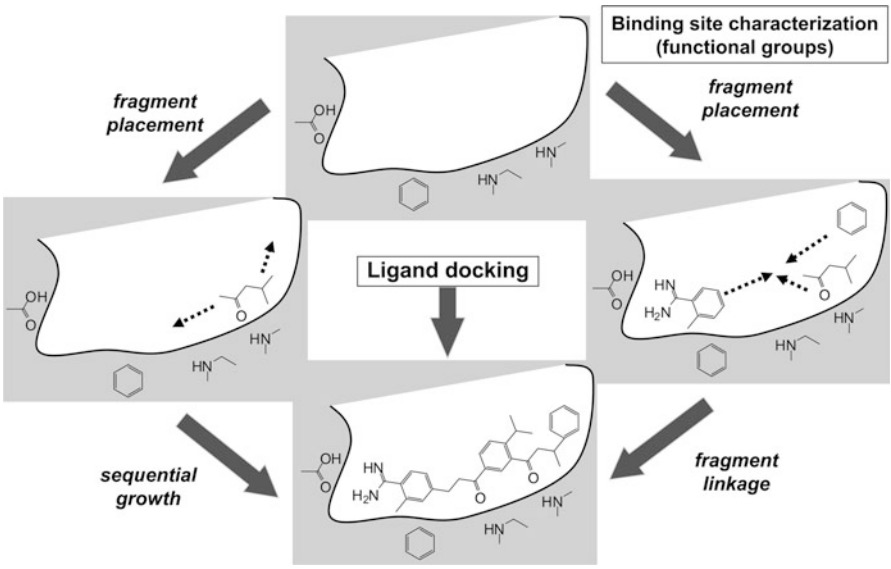


Fig. 20.8 Schematic diagram showing ligand- and fragment-based virtual screening strategies

cavities (Fig. 20.8). Docked fragments can then evolve virtually (by adding different chemical decorations), be linked (to join two or more fragments that occupy different regions of the binding site), self-assemble (through direct bond formation between different reacting fragments), and/or be optimized to better fulfill drug-like properties. Since the binding affinity of compounds has been shown to grow approximately linearly with their molecular weight, the affinity of fragments and standard drugs are usually compared in terms of a metric called ligand efficiency index, which is defined as the binding free energy per heavy atom. Given that affinity/activity information related to ligand-target complexes is currently stored in several databases (*e.g.* BindingDB, ChEMBL, and PDBBind, among others), ligand efficiency indices have been proposed as useful variables to chart the chemico-biological space and visualize it in an atlas-like type of representation [45].

The binding pockets that are explored in most SBVS campaigns are relatively small cavities that are meant for natural agonists, substrates or cofactors in a limited number of macromolecular targets. These represent what are known as the "orthosteric" binding sites for drugs. However, other drugs (*e.g.* NNRTIs) exert their action by altering the receptor's conformation in such a way that it would affect the binding and/or functional properties of agonists or substrates ("allosteric modulators") whereas yet others are known to modulate or prevent some key protein-protein interactions.

Most protein-protein interfaces have been usually regarded as "undruggable" because conventional analysis reveals them as lacking well-defined binding pockets. However, it has been noted that a large percentage of the free energy of interaction arises from small regions designated as "hot spots". These facts supported the concept by which protein-protein interaction disruptors do not need

to mimic the entire protein binding surface but rather a smaller subset of key residues in a relatively compact region of one of the binding partners.

Different computational tools have been developed for exploring the properties of protein-protein interfaces and facilitate the identification of starting points for small-molecule design (*e.g.* PocketQuery [46], http://pocketquery.csb.pitt.edu/). More generally, "druggable" sites can be located using FTMap [47], which is an efficient fast Fourier transform correlation method that samples multiple positions of up to 16 different probes on a three-dimensional grid centered on a protein of interest. A simple energy expression then allows the identification of consensus sites where probes tend to cluster, giving rise to favorable interactions. These sites correlate with locations where high-affinity ligands are found in protein complexes. The method is sensitive to conformational changes in the protein and can be used to map affinity sites prior to docking. The FTMap algorithm is implemented in the public domain Web server http://ftmap.bu.edu/.

20.4.3 Structure-Activity Relationships

Finally, visual examination of the crystal structure of a single ligand-receptor complex is usually highly informative and can guide the design of new analogues that will display graded affinities toward the target. However, a quantitative assessment of these variations will generally be difficult because the observed free energy differences result from a subtle interplay of binding forces within the receptor binding site that normally take place in the face of competition with water molecules. An approach that has been shown to be useful in these cases is to model the whole set of complexes and decompose the ligand-receptor interaction energies into a number of van der Waals and electrostatic contributions emanating from individual receptor amino acids, together with additional terms representing the desolvation of both binding partners. The resulting matrix of energy terms can then be projected onto a small number of orthogonal "latent variables" or principal components using partial least squares in a way similar to that used in comparative molecular field analysis (CoMFA). At the end of the procedure, those pair-wise interactions between the ligands and individual protein residues that are predictive of binding affinity or activity are selected and weighted according to their importance in the model. Since its inception, this chemometric method, termed comparative binding energy (COMBINE) analysis, has been successfully applied to a variety of biologically relevant targets including the protease and the RT of HIV-1 [48].

20.5 Case Studies in Structure-Based Antiviral Drug Development

During the last three decades we have witnessed an unprecedented effort to discover new antiviral drugs, most of them directed towards HIV, and more recently towards HCV. Many of those studies have benefited from the availability

of high-resolution crystal structures. We have selected a few case studies to illustrate how structure-based design has contributed to antiviral drug discovery. In this section we describe two successful examples that include oseltamivir and zanamivir acting on influenza virus neuraminidase, and mutation-resilient inhibitors of HIV-1 protease (*e.g.* darunavir). In the next section (20.6) we describe current experimental efforts in structure-based design, applied to inhibitors acting on the assembly or disassembly of viral capsids.

20.5.1 Neuraminidase Inhibitors Against the Threat of Influenza Pandemics

Although the crystal structure of influenza virus neuraminidase (also known as sialidase) was determined in 1983 [49], its resolution was relatively low (approx. 3.0 Å) and therefore, of limited value in structure-based drug design. Initial efforts to design neuraminidase inhibitors were based on substrate-like compounds such as *N*-acetylneuraminic acid (Neu5Ac) derivatives (Fig. 20.9). One of them, the 2-deoxy-α-Neu5Ac demonstrated good activity in mouse model experiments. The improvement in resolution of the neuraminidase X-ray structures obtained in complex with Neu5Ac or with the derivative Neu5Ac2en facilitated the design of more potent inhibitors. These studies revealed that sialic acid adopts a quite deformed conformation when bound to neuraminidase, due to strong ionic interactions of the sialic acid carboxyl moiety with three arginine residues of the enzyme (Arg118, Arg192 and Arg371). Starting with Neu5Ac2en, the substitution of the C-4 hydroxyl group with a guanidinyl functionality rendered the 4-deoxy-4-guanidino-Neu5Ac2en (zanamivir) that showed significantly improved affinity promoted by interactions between the guanidinyl moiety and the neuraminidase residues Glu119 and Glu227 [50]. Zanamivir was found to be highly selective for influenza virus sialidase. It was approved in 1999 as the first neuraminidase-targeting anti-influenza drug (Fig. 20.9) even though it can be administered only as a powder for oral inhalation.

Oseltamivir was developed as a result of efforts to develop neuraminidase inhibitors based on non-carbohydrate templates [50]. In this context, a cyclohexene ring was considered as a transition-state mimic, and amenable to chemical modifications to optimize its biological activity. X-ray crystallographic studies of sialic acid and related analogues bound to neuraminidase (Fig. 20.9) demonstrated that the C7 hydroxyl group of the glycerol side chain does not interact with amino acid residues in the sialidase active site. Replacement of the glycerol moiety to improve the lipophilicity of those inhibitors was carried out even though the hydroxyl groups at C8 and C9 were involved in interactions with Glu276 and Arg224. A 3-pentyl ether side chain at position C7 gave optimal results and led to the development of oseltamivir. Oseltamivir (originally known as GS 4104) is an ethyl ester prodrug of GS 4071 that can be administered orally as capsules or as a

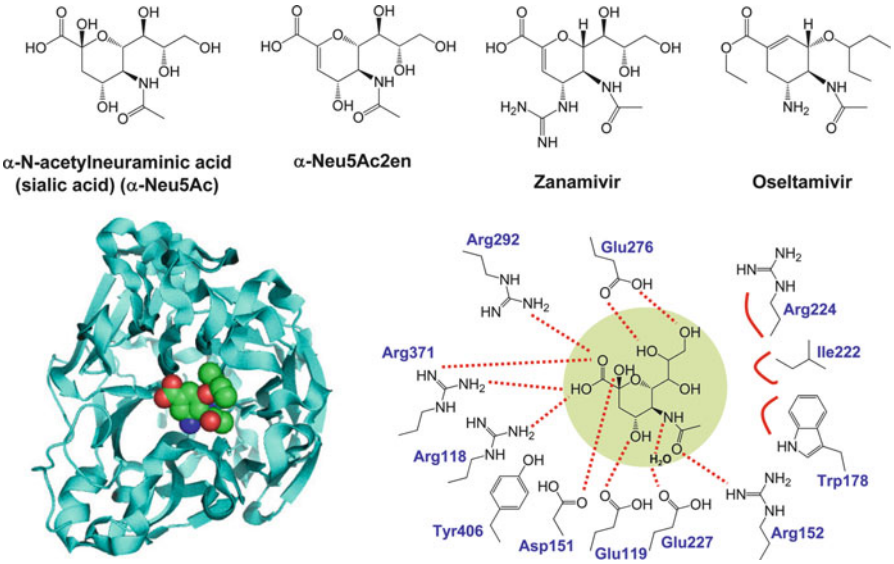


Fig. 20.9 Influenza virus neuraminidase inhibitors and structural basis of their mechanism of action. The lower part shows the structure of a neuraminidase monomer with a molecule of GS 4071 (the oseltamivir prodrug). C, O and N atoms in the inhibitor are shown in *green*, *red* and *purple*, respectively (shown structure was taken from the PDB file 2HT8). The scheme on the right shows hydrogen bonding network (*red dotted lines*) and hydrophobic pockets (*continuous red line*) in the sialic acid binding site of influenza virus neuraminidase (Based on the scheme published in Ref. [50])

suspension, and is then converted to the active form *in vivo* by the action of endogenous esterases. Crystal structures of influenza virus neuraminidase bound to GS 4071 showed a rearrangement of the side-chain of Glu276 that now interacts with Arg228, generating a larger hydrophobic area in this domain. The amino acid substitution H274Y decreases neuraminidase binding affinity for oseltamivir and is the major mutation involved in resistance to this antiviral drug. Crystal structures of oseltamivir-resistant mutant neuraminidases have shown that the presence of Tyr instead of His274 affects the positioning of Glu276 and disrupts the hydrophobic pocket that accommodates the 3-pentyl ether side chain of oseltamivir [35, 51].

20.5.2 Development of Darunavir to Combat Resistance to HIV-1 Protease Inhibitors

The HIV-1 protease is a homodimer that contains two catalytic aspartic acid residues (one in each subunit) in the active site (Fig. 20.5). The Asp25 residues share an acidic proton and interact with a water molecule in the absence of substrate or inhibitor. The hydrolysis mechanism involves activation of a water molecule by these aspartates and nucleophilic attack of the water oxygen to the amide carbonyl

of the bond to be cleaved between P1 and P1′. Breakdown of the resulting tetrahedral intermediate by successive proton transfers leads to the amino and carboxylate products. The scissile bond of the peptide substrate is in close proximity of the active site. Upon binding, flexible flaps (one per monomer) that adopt an open conformation in the apoenzyme fold down over the substrates. The peptide substrates contain at least seven residues extending from P4 to P3′, where the cleavage site lies between P1 and P1′. The side chains of the substrate lie in the protease subsites S4 to S3′. Replacement of the hydrolyzable peptide linkage by a non-cleavable bond was a basic principle for the development of peptidomimetic inhibitors that culminated with the approval of saquinavir [52].

Saquinavir and other related protease inhibitors have serious drawbacks that include severe side effects and toxicities, poor bioavailability (high doses required to maintain their therapeutic effect), costly synthesis and the rapid emergence of drug resistance. Extensive analysis of protein-ligand X-ray structures of wild-type and mutant HIV-1 proteases showed that there were minimal alterations in the active site backbone conformation of proteases resistant to inhibitors. Therefore, drugs effective on resistant HIV-1 strains should maintain extensive hydrogen-bonding interactions with the protein backbone of the wild-type enzyme.

The replacement of the asparagine moiety at the P2 position in saquinavir with 3*R*-tetrahydrofuranyl glycine resulted in a more potent inhibitor (Fig. 20.10). Other studies leading to the development of fosamprenavir/amprenavir showed an improvement in antiviral activity by introducing the 3*R*-tetrahydrofuranyl urethane moiety onto a hydroxyethylamine sulfonamide isostere. However, crystal structures revealed that 3*R*-tetrahydrofuranyl urethane-bearing inhibitors had weak hydrogen bonding potential with main chain atoms of residues Asp29 and Asp30. Therefore, a bicyclic tetrahydrofuran (bis-THF) ligand was designed and developed to improve hydrogen bonding with both aspartic acid residues. Inhibitors incorporating a P2 bis-THF and a *p*-methoxysulfonamide as the P2′ ligand (*e.g.* TMC-126) were remarkably potent in enzyme inhibition assays and displayed strong antiviral activity in cell culture. Darunavir was similar to TMC-126 but contained *p*-aminosulfonamide instead of *p*-methoxysulfonamide at the P2′ position. The P2′-amine group provided more favorable pharmacokinetic properties compared to the P2′-methoxy group. Darunavir-bound X-ray crystal structures showed that both P2 and P2′ moieties are involved in extensive hydrogen bonding with the protein backbone. These tight interactions were consistently observed with mutant proteases and might be responsible for the unusually high resistance profile of darunavir. Detailed reviews describing the design and development of darunavir have been published [53].

20.6 Viral Capsids as Targets of Antiviral Intervention

Remarkable examples of structure-based design of preclinical drugs binding to the mature or immature viral capsid or interfering with their assembly or conformational transitions required for genome uncoating have been described for picornaviruses, HBV and, more extensively, for HIV-1.

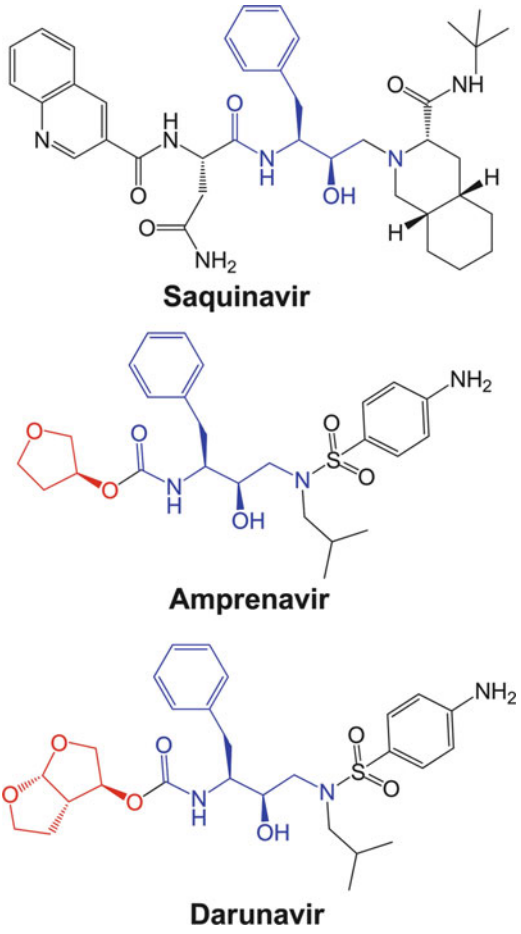


Fig. 20.10 Chemical structures of saquinavir, amprenavir and darunavir

20.6.1 Picornaviruses

The *Picornaviridae* family includes enteroviruses and rhinoviruses which are important human pathogens. Picornaviruses are non-enveloped, positive-stranded RNA viruses with an icosahedral capsid. The availability of crystal structures of poliovirus and rhinovirus (see Chaps. 4 and 10) led to the discovery of compounds that could bind to a hydrophobic pocket in the capsid protein VP1. Those molecules, known as WIN compounds (for Winthrop, the firm that developed the drugs), mimic natural factors that bind to the VP1 pocket and stabilize the capsid in a metastable conformation. During the course of a normal infection by poliovirus or rhinovirus, virion binding to the cell receptor allows for the displacement of the pocket-bound natural factors. This event triggers a global rearrangement of the capsid that is required for the release of the viral genome into the host cell cytoplasm [54]. WIN compounds and other related drugs bind to the pocket with

higher affinity than the natural factors and interfere with the receptor-induced conformational change by preventing either receptor binding or the receptor-induced rearrangement that is required for genome uncoating. In a long series of trend-setting studies, structure-function relationship analyses based on the determination of crystal structures of capsid-bound compounds, docking studies, genetic modifications of the virus particles and chemical modifications of the antiviral compounds led to the development of derivatives with improved antiviral activity.

These derivatives included pleconaril (WIN-63843) for the treatment of rhinovirus-induced exacerbation of pre-existing asthma or chronic obstructive pulmonary disease [55], BTA-798 for the enterovirus-induced sepsis syndrome [56], and V-073 with potential applications in the treatment of poliovirus infections [57]. Pleconaril binds into the hydrophobic pocket and stabilizes the metastable capsid conformation in a way that results in inhibition of both viral adsorption to cell membranes and RNA uncoating [54]. Natural resistance to this drug is largely confined to VP1 residue 191, which is valine in the vast majority of susceptible viruses and leucine in all resistant serotypes. Full resistance is achieved by an additional Tyr152 to Phe mutation [58].

20.6.2 Hepatitis B Virus

In HBV, the capsid is composed of 120 copies of a dimeric structural protein (known as CA protein) that assemble around an RNA-reverse transcriptase complex that will transcribe the RNA pregenome to the mature DNA genome. Ten years ago the fluorescent probe 5,5′-bis[8-(phenylamino)-1-naphtalenesulfonate] was shown to disturb the geometry of intersubunit contacts and to act both as a noncompetitive inhibitor of capsid assembly in HBV and as an assembly 'misdirector' that facilitates the formation of noncapsid polymers [59]. Assembly misdirection has also been shown to occur upon exposure to activator heteroaryldihydropyrimidines such as HAP1 [28, 60] and BAY 41–4109 [61].

20.6.3 Human Immunodeficiency Virus

In the case of HIV-1, we have discussed above the role of the viral protease in the release of the capsid protein (CA) from the precursor Gag polyprotein, and how protease inhibitors can prevent the formation of the mature capsid. Another compound that interferes with the maturation process is bevirimat (3-*O*-(3′,3′-dimethylsuccinyl)betulinic acid, also known as PA-457). This compound is a derivative of betulinic acid and interferes with the processing of the Gag polyprotein at the CA-SP1 cleavage site (*i.e.* between the C-terminal of the capsid protein and a spacer peptide located between CA and the nucleocapsid protein). This inhibitory action is independent of the protease activity. Tomographic reconstructions suggest

that the CA-SP1 of Gag forms a helical bundle that serves to anchor the carboxy-terminal domain (CTD) of CA below the plane of the N-terminal domain (NTD). In the absence of CA-SP1 cleavage, mature cores cannot be formed because the immature particles cannot disassemble (for reviews, see [30, 62]). Virions produced from bevirimat-treated cells display aberrant core morphologies. A similar mechanism of action has been recently described for a pyridine-based compound (PF-46396) whose inhibitory activity results in the accumulation of CA/SP1 (p25) precursor proteins and the blockade of viral core particle maturation [63].

Assembly of the HIV-1 capsid depends on CA-CA interactions involving structural domains NTD and CTD (Fig. 20.11). The peptides CAI and CAC1 and derivatives, as well as small molecules such as α-hydroxyglycineamide (a C-amidated tripeptide), benzodiazepine-based compounds and CAP-1 have been identified as inhibitors of HIV-1 capsid assembly [29, 30, 65]. CAP-1 contains a urea-based scaffold and two substituted aromatic groups at the molecule ends, and acts in the late stage of the viral life cycle. When added to preformed HIV-1 particles, the drug has no inhibitory effect on HIV-1 infection. The CAI peptide was identified in a screen using a random peptide library. The CAC1 peptide was rationally designed based on structure and thermodynamic information to include most of the residues critically involved in the CTD-CTD dimerization interface that participates in mature HIV-1 capsid assembly [65]. CAP-1, CAI, CAC1 and their combinations and/or derivatives were able to partially inhibit HIV-1 infectivity in cultured cells.

One of the limitations of CAI and other peptide inhibitors of capsid assembly is their inability to penetrate cells. Chemical modifications consisting of covalent crosslinks between amino acids found in those inhibitors led to the development of conformationally restricted peptides that showed increased affinity for monomeric CTD and improved delivery to target cells. Examples of those "stapled" peptides are NYAD-1 (derivative of CAI) and NYAD-201 [65]. NYAD-1 inhibits the formation of both immature and mature virions. At low micromolar concentrations, this drug interferes with early events in the viral life cycle. Despite their higher solubility and low toxicity, further development of HIV-1 CA-derived natural or stapled peptides is limited by their relatively modest binding affinities [30, 66]. Also, problems related with the general metabolic instability of peptides may need to be addressed by using peptidomimetic compounds.

A high-throughput antiviral screening campaign led to the identification of a cluster of structurally related compounds that inhibited HIV-1 replication at submicromolar concentrations by targeting a post-entry stage prior to reverse transcription. Among them, PF-3450074 (or PF74) triggers premature HIV-1 uncoating in target cells [67]. Interestingly, point mutations in CA that stabilize the HIV-1 core conferred strong resistance without influencing inhibitor binding. Structural studies of the complex of the CA N-terminal domain and PF74 (PDB entry 2XDE) revealed a new binding site, different from those described for CAP-1 (PDB entry 2PXR) or CAI (PDB entry 2BUO) (Fig. 20.11), and several *in silico* screened inhibitors [68]. PF74 increased the rate of HIV-1 CA multimerization *in vitro,* while CAI and CAP-1 decreased this rate under the same assay conditions. In a model of an assembled CA hexamer in complex with PF74, an indole group

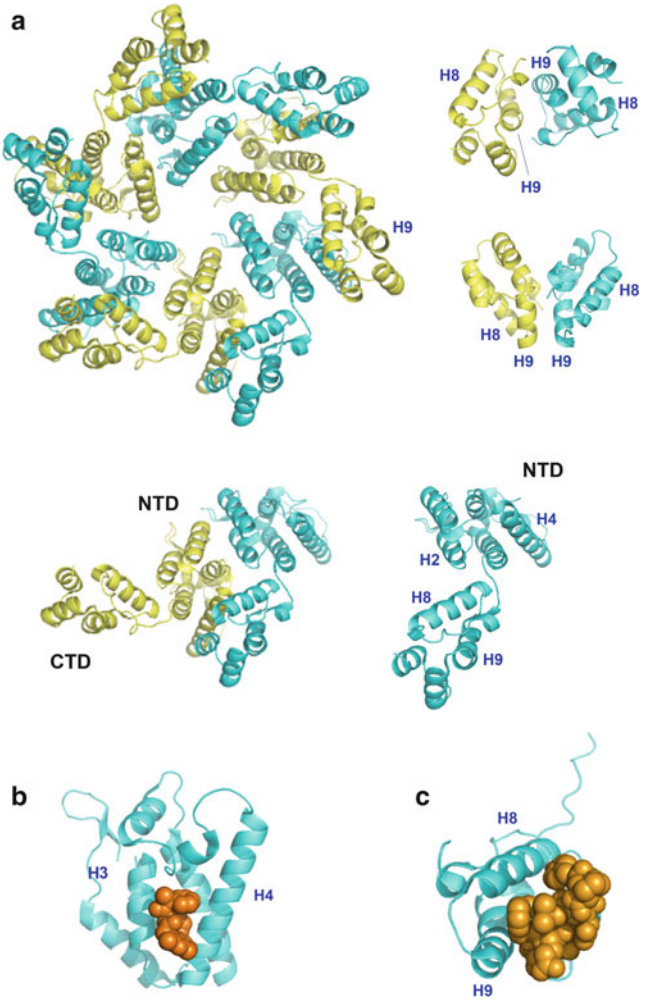


Fig. 20.11 HIV-1 capsid structure as a target of specific inhibitors. (**a**) Structural model of the HIV-1 CA hexamer [64] (coordinates taken from PDB entry 3GV2). The cartoons below show the structures of a CA dimer and a monomer, indicating the location of the structural domains NTD and CTD. Helices H2, H4 and H8 contribute to the NTD-CTD and NTD-NTD interfaces. CTD-CTD interactions are important for the formation of the capsid lattice by connecting hexameric rings. CTD dimerization is largely dependent on interactions between H9 helices (CTD dimers shown on the right are models obtained with data from the PDB entry 1A43). (**b**) Structure of a monomeric CA NTD bound to PF74 (PDB entry 2XDE). (**c**) Structure of a monomeric CA CTD bound to CAI (PDB entry 2BUO). All figures were produced using PyMol (http://www.pymol.org/)

protruding from the N-terminal domain was found to localize to the interface between capsid monomers suggesting both early and late stage effects on viral propagation. The N-terminal domain of CA is also the target of novel benzodiazepines that prevent virus release or benzimidazoles that inhibit the

formation of the mature capsid [69]. These compounds bind to a different site as compared with PF74 and have no effects at the early steps of viral replication. Benzimidazol inhibitors are expected to interfere with the formation of CA_{NTD}/CA_{CTD} interfaces in the CA hexamer [69].

20.7 Perspectives and Conclusions

Despite the recent success in developing new antiviral drugs, particularly against HIV and HCV, there are still important challenges ahead. Drug resistance constitutes a frequent obstacle that limits the efficacy of antiviral treatments. As learnt from the HIV experience, monotherapies should be considered suboptimal therapies and combination therapies are expected to prevail in the future treatment of viral infections. Good combination therapies rely on the exploitation of multiple targets and/or mechanisms of action. A better knowledge of the biological processes involved in specific viral infections, including interaction with specific host factors, is required to develop efficient screening methods and to identify viable targets. A wealth of information derived from massive sequencing techniques applied to humans and pathogens will be probably helpful in this process. Drug efficacy, toxicity and cost are also issues that need to be considered in the development of future antiviral drugs.

From a technical point of view, high-throughput X-ray crystallography and structural analysis, computational modeling and better synthetic methods will provide tools to accelerate drug discovery, once the target mechanism has been validated. The rapidly increasing availability of high-resolution structures of viral particles, and substantial advances in our understanding of their morphogenesis and conformational dynamics are opening new avenues for the design of novel antiviral agents. Although the odds of finding effective inhibitors of novel targets have probably increased, the outcome still remains unpredictable.

Although HIV and HCV will probably continue to be major actors in the drug development field, other viral pathogens such as influenza virus and flaviviruses (*e.g.* dengue virus) will probably gain attention in the coming years due to their general impact on human health.

Acknowledgements Work at CBMSO is supported in part by grants of the Spanish Ministry of Economy and Competitiveness (BIO2010/15542) and the General Directorate for Pharmacy and Medical Products (Ministry of Health, Social Services and Equality) (EC11-025), and an institutional grant from the Fundación Ramón Areces. Continued support to F.G. from the Spanish Comisión Interministerial de Ciencia y Tecnología (SAF2009-13914-C02-02) and Comunidad de Madrid (S-BIO/0214/2006 and S2010-BMD-2457) is gratefully acknowledged.

References

1. De Clercq E (2004) Antivirals and antiviral strategies. Nat Rev Microbiol 2:704–720
2. De Clercq E (2009) Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. Int J Antimicrob Agents 33:307–320
3. Menéndez-Arias L (2010) Molecular basis of human immunodeficiency virus drug resistance: an update. Antiviral Res 85:210–213
4. Menéndez-Arias L (2009) Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. Viruses 1:1137–1165
5. Hamre D, Brownlee KA, Donovick R (1951) Studies on the chemotherapy of vaccinia virus. II The activity of some thiosemicarbazones. J Immunol 67:305–312
6. Bauer DJ, St. Vincent L, Kempe CH, Downie AW (1963) Prophylactic treatment of smallpox contacts with N-methylisatin β-thiosemicarbazone (compound 33T57, Marboran). Lancet 2:494–496
7. Prusoff WH (1959) Synthesis and biological activities of iododeoxyuridine, an analog of thymidine. Biochim Biophys Acta 32:295–296
8. Jordan AM, Roughley SD (2009) Drug discovery chemistry: a primer for the non-specialist. Drug Discov Today 14:731–744
9. De Clercq E, Holý A (2005) Acyclic nucleoside phosphonates: a key class of antiviral drugs. Nat Rev Drug Discov 4:928–940
10. Menéndez-Arias L (2008) Mechanisms of resistance to nucleoside analogue inhibitors of HIV-1 reverse transcriptase. Virus Res 134:124–146
11. Menéndez-Arias L, Betancor G, Matamoros T (2011) HIV-1 reverse transcriptase connection subdomain mutations involved in resistance to approved non-nucleoside inhibitors. Antiviral Res 92:139–149
12. De Corte BL (2005) From 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-*jk*](1,4)benzodiazepin-2(1*H*)-one (TIBO) to etravirine (TMC125): Fifteen years of research on non-nucleoside inhibitors of HIV-1 reverse transcriptase. J Med Chem 48:1689–1696
13. Janssen PAJ, Lewi PJ, Arnold E et al (2005) In search of a novel anti-HIV drug: multidisciplinary coordination in the discovery of 4-[[4-[[-[(1*E*)-2-cyanoethenyl]-2, 6-dimethylphenyl] amino]-2-pyrimidinyl]amino]benzonitrile (R278474, rilpivirine). J Med Chem 48:1901–1909
14. Nassal M (2008) Hepatitis B viruses: reverse transcription a different way. Virus Res 134:235–249
15. Watkins WJ, Ray AS, Chong LS (2010) HCV NS5B polymerase inhibitors. Curr Opin Drug Discov Devel 13:441–465
16. Mayhoub AS (2012) Hepatitis C RNA-dependent RNA polymerase inhibitors: a review of structure-activity and resistance relationships; different scaffolds and mutations. Bioorg Med Chem 20:3150–3161
17. Sofia MJ, Bao D, Chang W et al (2010) Discovery of a β-D-2′-deoxy-2′-α-fluoro-2′-β-*C*-methyluridine nucleotide prodrug (PSI-7977) for the treatment of hepatitis C virus. J Med Chem 53:7202–7218
18. De Clercq E (2007) The design of drugs for HIV and HCV. Nat Rev Drug Discov 6:1001–1018
19. Gao M, Nettles RE, Belema M et al (2010) Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature 465:96–100
20. Lemm JA, O'Boyle D 2nd, Liu M et al (2010) Identification of hepatitis C virus NS5A inhibitors. J Virol 84:482–491
21. Coelmont L, Hanoulle X, Chatterji U et al (2010) DEB025 (Alisporivir) inhibits hepatitis C virus replication by preventing a cyclophilin A induced cis-trans isomerisation in domain II of NS5A. PLoS One 5:e13687
22. Hazuda D, Iwamoto M, Wenning L (2008) Emerging pharmacology: inhibitors of human immunodeficiency virus integration. Annu Rev Pharmacol Toxicol 49:377–394

23. Serrao E, Odde S, Ramkumar K, Neamati N (2009) Raltegravir, elvitegravir, and metoogravir: the birth of "me-too" HIV-1 integrase inhibitors. Retrovirology 6:25
24. Cherepanov P, Maertens GN, Hare S (2011) Structural insights into the retroviral DNA integration apparatus. Curr Opin Struct Biol 21:249–256
25. Njoroge FG, Chen KX, Shih NY, Piwinski JJ (2008) Challenges in modern drug discovery: a case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. Acc Chem Res 41:50–59
26. Kwong AD, Kauffman RS, Hurter P, Mueller P (2011) Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. Nat Biotechnol 29:993–1003
27. Hacker HJ, Deres K, Mildenberger M, Schröder CH (2003) Antivirals interacting with hepatitis B virus core protein and core mutations may misdirect capsid assembly in a similar fashion. Biochem Pharmacol 66:2273–2279
28. Stray SJ, Bourne CR, Punna S et al (2005) A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. Proc Natl Acad Sci USA 102:8138–8143
29. Neira JL (2009) The capsid protein of human immunodeficiency virus: designing inhibitors of capsid assembly. FEBS J 276:6110–6117
30. Prevelige PE Jr (2011) New approaches for antiviral targeting of HIV assembly. J Mol Biol 410:634–640
31. Gamblin SJ, Skehel JJ (2010) Influenza hemagglutinin and neuraminidase membrane glycoproteins. J Biol Chem 285:28403–28409
32. Menéndez-Arias L, Esté JA (2004) HIV-resistance to viral entry inhibitors. Curr Pharm Des 10:1845–1860
33. Madani N, Schön A, Princiotto AM et al (2008) Small-molecule CD4 mimics interact with a highly conserved pocket on HIV-1 gp120. Structure 16:1689–1701
34. Curreli F, Choudhury S, Pyatkin I et al (2012) Design, synthesis, and antiviral activity of entry inhibitors that target the CD4-binding site of HIV-1. J Med Chem 55:4764–4775
35. Colman PM (2009) New antivirals and drug resistance. Annu Rev Biochem 78:95–118
36. Roymans D, De Bondt HL, Arnoult E et al (2010) Binding of a potent small-molecule inhibitor of six-helix bundle formation requires interactions with both heptad-repeats of the RSV fusion protein. Proc Natl Acad Sci USA 107:308–313
37. Chen GS, Chern J-W (2007) Computer-aided drug design. In: Hwang Z (ed) Drug discovery research: new frontiers in the post-genomic era. Wiley, Hoboken, pp 89–107
38. McInnes C (2007) Virtual screening strategies in drug discovery. Curr Opin Chem Biol 11:494–502
39. Irwin JJ, Shoichet BK (2005) ZINC–a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182
40. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
41. Horvath D (2011) Pharmacophore-based virtual screening. Methods Mol Biol 672:261–298
42. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3:935–949
43. Guido RV, Oliva G, Andricopulo AD (2008) Virtual screening and its integration with modern drug design technologies. Curr Med Chem 15:37–46
44. Sun C, Petros AM, Hajduk PJ (2011) Fragment-based lead discovery: challenges and opportunities. J Comput Aided Mol Des 25:607–610
45. Abad-Zapatero C, Perišić O, Wass J et al (2010) Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. Drug Discov Today 15:804–811
46. Koes DR, Camacho CJ (2012) PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. Nucleic Acids Res 40:W387–W392

47. Hall DR, Ngan CH, Zerbe BS, Kozakov D, Vajda S (2012) Hot spot analysis for driving the development of hits into leads in fragment-based drug discovery. J Chem Inf Model 52:199–209
48. Rodríguez-Barrios F, Gago F (2004) Chemometrical identification of mutations in HIV-1 reverse transcriptase conferring resistance or enhanced sensitivity to arylsulfonylbenzonitriles. J Am Chem Soc 126:2718–2719
49. Varghese JN, Laver WG, Colman PM (1983) Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature 303:35–40
50. Von Itzstein M (2007) The war against influenza: discovery and development of sialidase inhibitors. Nat Rev Drug Discov 6:967–974
51. Collins PJ, Haire LF, Lin YP et al (2008) Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. Nature 453:1258–1261
52. Roberts NA, Martin JA, Kinchington D et al (1990) Rational design of peptide-based HIV proteinase inhibitors. Science 248:358–361
53. Ghosh AK, Anderson DD, Weber IT, Mitsuya H (2012) Enhancing protein backbone binding – a fruitful concept for combating drug-resistant HIV. Angew Chem Int Ed Engl 51:1778–1802
54. Patick AK (2006) Rhinovirus chemotherapy. Antiviral Res 71:391–396
55. Ledford RM, Collett MS, Pevear DC (2005) Insights into the genetic basis for natural phenotypic resistance of human rhinoviruses to pleconaril. Antiviral Res 68:135–138
56. Thibaut HJ, De Palma AM, Neyts J (2012) Combating enterovirus replication: state-of-the-art on antiviral research. Biochem Pharmacol 83:185–192
57. Oberste MS, Moore D, Anderson B et al (2009) *In vitro* antiviral activity of V-073 against polioviruses. Antimicrob Agents Chemother 53:4501–4503
58. Kistler AL, Webster DR, Rouskin S et al (2007) Genome-wide diversity and selective pressure in the human rhinovirus. Virol J 4:40
59. Zlotnick A, Ceres P, Singh S, Johnson JM (2002) A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. J Virol 76:4848–4854
60. Bourne CR, Finn MG, Zlotnick A (2006) Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. J Virol 80:11055–11061
61. Stray SJ, Zlotnick A (2006) BAY 41–4109 has multiple effects on hepatitis B virus capsid assembly. J Mol Recognit 19:542–548
62. Adamson CS, Freed EO (2008) Recent progress in antiretrovirals – lessons from resistance. Drug Discov Today 13:424–432
63. Blair WS, Cao J, Fok-Seang J et al (2009) New small-molecule inhibitor class targeting human immunodeficiency virus type 1 virion maturation. Antimicrob Agents Chemother 53:5080–5087
64. Pornillos O, Ganser-Pornillos BK, Kelly BN et al (2009) X-ray structures of the hexameric building block of the HIV capsid. Cell 137:1282–1292
65. Bocanegra R, Rodríguez-Huete A, Fuertes MA et al (2012) Molecular recognition in the human immunodeficiency virus capsid and antiviral design. Virus Res 169:388–410
66. Zhang H, Curreli F, Zhang X et al (2011) Antiviral activity of α-helical stapled peptides designed from the HIV-1 capsid dimerization domain. Retrovirology 8:28
67. Shi J, Zhou J, Shah VB et al (2011) Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. J Virol 85:542–549
68. Blair WS, Pickford C, Irving SL et al (2010) HIV capsid is a tractable target for small molecule therapeutic intervention. PLoS Pathog 6:e1001220
69. Lemke CT, Titolo S, von Schwedler U et al (2012) Distinct effects of two HIV-1 capsid assembly inhibitor families that bind the same site within the N-terminal domain of the viral CA protein. J Virol 86:6643–6655

Further Reading

Kazmierski WM (ed) (2011) Antiviral drugs: from basic discovery through clinical trials. Wiley, Hoboken

LaFemina RL (ed) (2009) Antiviral research: strategies in antiviral drug discovery. ASM Press, Washington

Young DC (2009) Computational drug design: a guide for computational and medicinal chemists. Wiley, Hoboken

A collection of reviews written by Dr. Eric De Clercq and published in *Medicinal Research Reviews* between 2008 and 2011 provide a nice summary on the design and development of many antiviral drugs, from a historical perspective and providing relevant chemical structures. References for these articles are

De Clercq E (2008) The discovery of antiviral agents: ten different compounds, ten different stories. Med Res Rev 28:929–953

De Clercq E (2009) Antiviral drug discovery: ten more compounds, and ten more stories (part B). Med Res Rev 29:571–610

De Clercq E (2009) Another ten stories in antiviral drug discovery (part C): "old" and "new" antivirals, strategies, and perspectives. Med Res Rev 29:611–645

De Clercq E (2010) Yet another ten stories on antiviral drug discovery (part D): paradigms, paradoxes, and paraductions. Med Res Rev 30:667–707

De Clercq E (2011) The next ten stories on antiviral drug discovery (part E): advents, advances, and adventures. Med Res Rev 31:118–160

Also especially recommended for further reading are references [1, 3, 18, 35] *listed above.*

Chapter 21
Design of Novel Vaccines Based on Virus-Like Particles or Chimeric Virions

Juan Bárcena and Esther Blanco

Abstract Virus-like particles (VLPs) are formed by viral structural proteins that, when overexpressed, spontaneously self-assemble into particles that are antigenically indistinguishable from infectious virus or subviral particles. VLPs are appealing as vaccine candidates because their inherent properties (*i.e.*, virus-sized, multimeric antigens, highly organised and repetitive structure, not infectious) are suitable for the induction of safe and efficient humoral and cellular immune responses. VLP-based vaccines have already been licensed for human and veterinary use, and many more vaccine candidates are currently in late stages of evaluation. Moreover, the development of VLPs as platforms for foreign antigen display has further broadened their potential applicability both as prophylactic and therapeutic vaccines. This chapter provides an overview on the design and use of VLPs for the development of new generation vaccines.

Keywords Virus • Virus-like particles • VLPs • Chimeric VLPs • Viral nanoparticles • Capsid structure • Structural proteins • Vaccines • Immune response • Multimeric presentation • Antigen display • Epitopes • B-cell • T-cell

Abbreviations

ANTXR2	Anthrax toxin receptor 2
APC	Antigen-presenting cell
BTV	Bluetongue virus
CpG-ODN	CpG-containing oligodeoxynucleotide
CPMV	Cowpea mosaic virus
CPV	Canine parvovirus

J. Bárcena (✉) • E. Blanco
Centro de Investigación en Sanidad Animal (INIA), Valdeolmos, 28130 Madrid, Spain
e-mail: barcena@inia.es

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_21,

cryo-EM	Cryo-electron microscopy
CTL	Cytotoxic T cell
DC	Dendritic cell
FHV	Flock house virus
GFP	Green fluorescent protein
HBcAg	Hepatitis B core antigen
HBsAg	Hepatitis B small surface antigen
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HRV-14	Human rhinovirus 14
IBDV	Infectious bursal disease virus
MEV	Mink enteritis virus
MHC	Major histocompatibility complex
MIR	Major immunodominant region
PA	Protective antigen
PAMP	Pathogen-associated molecular pattern
PCV2	Porcine circovirus type 2
PRR	Pattern-recognition receptor
TLR	Toll-like receptor
VLP	Virus-like particle

21.1 Introduction

Vaccination is one of the most important and cost-effective methods of preventing infectious diseases. To date, no other method in human or veterinary medicine has had such an impact in reducing morbidity and mortality and increasing the overall well-being of humans and animals [1–5]. Life-threatening and devastating diseases, such as smallpox, have been eradicated by vaccination campaigns and the incidence of many others has been greatly reduced. Historically, this notable achievement has been brought about by an essentially empirical approach. Most of currently licensed vaccines are still produced by relatively simple methods of reducing the pathogen's virulence. These include biological attenuation of virulence, for example by repeated passage of viruses in cell-culture or embryonated eggs (live-attenuated vaccines), and chemical inactivation of whole microorganisms (inactivated or "killed" vaccines). Upon administration, these classical vaccines, especially the live-attenuated ones, frequently induce a spectrum of immune responses which closely resembles that elicited by the actual infectious agent: a potent B-cell mediated (or humoral) immune response plus a T-cell mediated (or cellular) immune response (Fig. 21.1), eventually conferring life-long protection of the vaccinated host against subsequent infection by the corresponding pathogen.

Nevertheless, in spite the successful development of numerous efficient classical vaccines their use presents relevant drawbacks regarding safety and other issues. There is a limited but present risk of reversion to a virulent phenotype *in vivo*, incomplete inactivation, or accidental pathogen release from vaccine manufacturing facilities, any of which can lead to spreading of the disease the vaccine was intended to prevent. Examples of this have been observed involving classical vaccines against poliovirus or foot-and- mouth disease virus [6, 7]. Live-attenuated vaccines are contraindicated for use in immunocompromised individuals or newborns. Besides, in the veterinary field, the use of classical vaccines do not usually allow for a clear-cut serological discrimination between vaccinated and pathogen-infected animals, a mandatory requirement to gain access to the international market of livestock products. These concerns have led many countries to a non-vaccination policy against important livestock diseases, that relies on slaughtering infected and contact herds, together with strict limitations on animal movements. Consequently, livestock populations are highly susceptible to such diseases, and there is a serious risk of reintroduction, as has happened repeatedly during the last years in several European countries, with diseases such as classical swine fever or foot-and-mouth disease. Another critical issue regarding the production of both human and animal classic viral vaccines is the need to develop a cell culture system susceptible of infection by the viral pathogen, that has not been always possible and/or advisable due, for example, to technical or safety reasons. Finally, the traditional strategies have shown their limits for the development of vaccines against organisms with complex pathogen-host interactions, that profit from undermining, evading, or misdirecting the host's immune response, such as human immunodeficiency virus (HIV). Hence, the focus of modern vaccinology is the development of alternative vaccine approaches circumventing the limitations of classical vaccines.

The advent of genetic engineering techniques and new biotechnological methods have enabled the development of synthetic or subunit vaccines, based on the use of isolated components of pathogens, such as recombinant immunogenic proteins or synthetic peptides, often portions of virus capsids. Subunit vaccines are considered an intrinsically safe approach as they are based on non-replicating immunogens and are, thus absolutely non-infectious. Importantly for veterinary use, they allow for discrimination between vaccinated and pathogen-infected animals (the so called marker vaccines). However, safety comes at a price. Vaccines based on isolated, soluble components of viruses and other pathogens are usually poorly immunogenic *per se*, requiring the administration of frequent and large doses of antigen. Moreover, they usually need to be formulated with potent immune-stimulating adjuvants, which may cause considerable side effects such as toxicity or pain.

Basic studies on virus structure and assembly (see Chaps. 2, 10 and 11) led to the experimental observation that many viral structural proteins have the intrinsic ability to self-assemble into virus-like particles (VLPs). These VLPs have not only been used in fundamental studies on capsid assembly, conformational stability and dynamics, and disassembly, but have also led to better immunological mimics of whole-virus particles compared to capsid subunits, resulting in improved effectiveness as vaccines, and leading to a renaissance in vaccine development. VLP

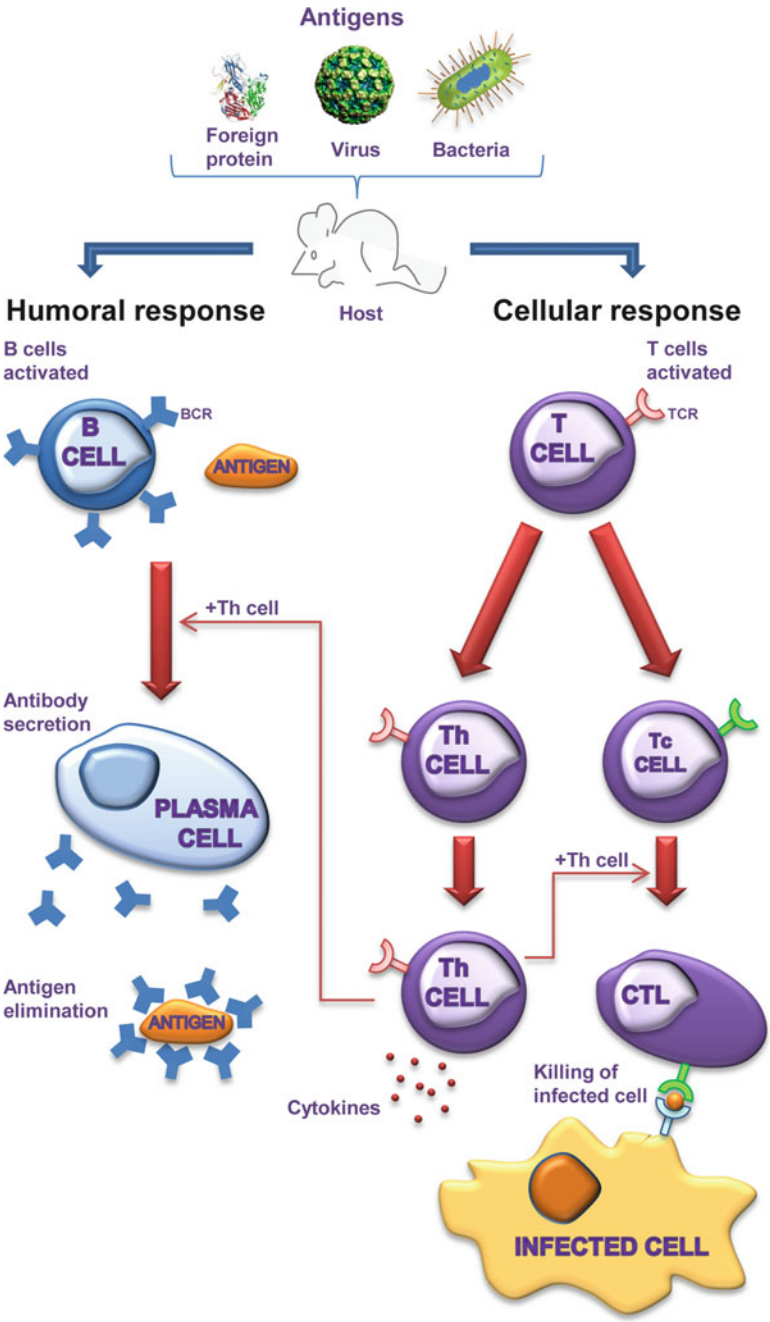


Fig. 21.1 The humoral and cell-mediated branches of the immune response. The humoral branch comprises lymphocytes of the B-cell lineage. Antibodies are the effector molecules produced by this response. The process begins with the interaction of B-cell receptors (*BCRs*) with antigens. Binding of the antigen promotes differentiation into antibody-secreting cells (plasma cells). The cell-mediated branch comprises lymphocytes of the T-cell lineage. T-helper cells (*Th cells*) and

based-vaccines combine many of the advantages of whole-virus-based and recombinant subunit vaccines, exhibiting a high safety profile. VLPs, produced using recombinant protein expression systems (thus avoiding the need to handle large amounts of pathogenic microorganisms), can stimulate strong B and T-cell immune responses and have been shown to exhibit self-adjuvanting abilities. These characteristics have made VLPs attractive stand-alone vaccine candidates for many viral diseases [8–13]. In addition, VLPs can also be used as platforms for the multimeric display of foreign antigens of interest, derived from viruses or other pathogens. This strategy couples delivery of the chosen antigen with the ability of VLPs to induce strong immune responses, thus acting as efficient adjuvants [14–17].

Until recently, vaccines were exclusively used to prevent infectious diseases (prophylactic vaccines). This situation is evolving rapidly and many efforts are currently focused on developing therapeutic vaccines to target diseases such as cancer, allergy and inflammatory autoimmunity, or even new indications like vaccines against nicotine dependence or hypertension [18–20]. In recent years, mechanisms underlying the outstanding immunogenicity of viruses have been elucidated, identifying important key parameters, like their size, particulate nature, their highly repetitive and ordered structures (see Chap. 2), or their ability to activate critical subsets of cells from the innate immune response, which in turn triggers appropriate conditioning of the adaptive immune response [21–23]. These features now serve as a paradigm for rational, virus structure-based vaccine design, and many new-generation tailor-made vaccines based on VLPs are attempting to harness these viral properties to target a wide array of human an animal diseases [23, 24]. This chapter provides an overview on the development and use of VLPs as vaccines for prophylactic and therapeutic applications.

21.2 Immunology of Vaccines, an Overview

What makes VLPs good vaccine candidates? In order to answer this question it is important to briefly review how vaccines induce a protective immune response.

The immune system has evolved to distinguish between noninfectious self and infectious nonself. Accordingly, it can mount strong immune responses to microbial infections but usually fails to respond to self molecules. Unravelling the mechanisms used by the immune system to discriminate self from nonself is critical for our understanding of diseases and our ability to rationally design vaccines.

The immune system is composed of the innate (nonspecific) and adaptive (specific) branches [21, 25]. The innate immune system represents the first line of

Fig. 21.1 (continued) cytotoxic T cells (*CTLs*) are the key effectors of this response. The process begins when T lymphocytes recognise antigens on the surface of antigen presenting cells (*APCs*). The Th cells recognise antigens bound to MHC class II molecules and produce cytokines that promote activation of other lymphocytes (both B and T cells). Once activated by Th cells, Tc cells differentiate into CTLs that can kill pathogen-infected cells

host defense against microbial infections. After a pathogen breaches the host's physical barriers such as mucosal surfaces or skin, the immune system mounts an early, innate immune response within minutes. Cells of the innate immune system, like dendritic cells (DCs) and macrophages, known as antigen-presenting cells (APCs), can recognise general characteristics of the pathogen and initiate a response. Pathogen-associated molecular patterns (PAMPs) are molecules or motifs that are common to many pathogens and generally not present in the host, such as lipopolysaccharide, which can be found on the cell wall of many bacterial species, double-stranded RNA or unmethylated CpG motifs, which are normally associated with virus infection. PAMPs can be recognised by Toll-like receptors (TLRs) and other pattern-recognition receptors (PRRs) which are present in host cells. The multivalent display and highly ordered structure exhibited by many pathogens, notably icosahedral virus surfaces that are essentially two-dimensional crystals, are also recognised as PAMPs, resulting in increased immunogenicity. PAMPs trigger innate immune sensing mechanisms and stimulate antigen uptake and processing by APCs. Subsequently, activated APCs present the antigens to cells of the adaptive immune response.

Adaptive immunity represents the second line of immunological defense and takes longer to develop, on the order of days. Once the antigen is delivered to the adaptive immune system and stimulates the proliferation (clonal selection and expansion) of antigen-specific effector cells, these cells begin eliminating the pathogen and infected cells. The adaptive response is also the basis for immunological memory, which is important for ensuring a fast and strong response when an infectious pathogen is encountered again (secondary infections). Vaccines use this process by introducing the immune system to a specific pathogen, stimulating the formation of specific memory cells and antibodies, which are then available to rapidly recognise the natural pathogen upon future exposures, resulting in its elimination and termination of the disease. The primary players of the adaptive immune response are B and T cells (lymphocytes) (Fig. 21.1). The key function of B cells is the production of antibodies. The role of T cells is more diverse ranging from lysis of infected cells, carried out by the cytotoxic T cells (CTLs), to secretion of cytokines, stimulation of B cells and other regulatory functions, performed by helper T cells (Th cells). The contribution of B cells *versus* T cells, and CTLs *versus* Th cells to resolution of infection varies from pathogen to pathogen and from primary to secondary responses [25]. Although it is difficult to generalise, antibodies are usually more important during secondary infections. In contrast, T cells are more important during many primary infections and, in particular, for the containment of chronic infections. Thus, for prophylactic vaccines, which aim at inducing effector mechanisms that protect from reinfection, the induction of neutralising antibodies is usually the most important factor. In contrast, for therapeutic vaccination against chronic infections, the emphasis is usually the stimulation of CTL responses. Hence, rational design of vaccines entails knowledge on the type of immune response to be evoked, and this in turn has structural implications for the design of new generation vaccines.

21.3 The Role of Size, Geometry and Molecular Patterns in Vaccine Design

The key properties of viruses that are responsible for eliciting potent immune responses and may be used as framework for vaccine design include their size, geometry, and the ability to induce innate immunity with appropriate conditioning of the adaptive immune responses [26].

The dimensions of vaccine antigens vary greatly [27–29]. The smallest (<10 nm) are protein or subunit antigen vaccines. Often, such antigens are formulated with adjuvants (such as alum and Freund's adjuvants) to form larger particles or aggregates. Supramolecular particulate antigens, such as VLPs and nanoparticles, are larger (20–200 nm). Small liposomes, such as virosomes (VLPs derived from viruses with lipid envelopes) are approximately 100–200 nm. Antigens presented in the context of microparticles, liposomes, water in oil emulsions, mineral salts and whole-cell vaccines are the largest (100 nm–20 μm).

The size of antigens is an important factor for their efficient uptake by APCs [26–29]. Particulate antigens, such as whole-cell vaccines, virosomes and VLPs, or antigens formulated in particulate adjuvants, such as liposomes and microparticles, have large surfaces with charge, hydrophobic or receptor-interacting properties, enabling better interaction with APCs. Thus, in contrast to small soluble protein antigens, pathogen sized particles and protein aggregates may efficiently be taken up by APCs. Following uptake, antigens are processed in the endosomal-lysosomal compartments of APCs, where they are degraded into peptides. These peptides are then loaded onto major histocompatibility complex (MHC) class II molecules and are subsequently transported to the cell surface for stimulation of Th cells (Fig. 21.2). Additionally, through the process termed cross-presentation, DCs (a special subset of APCs) have the ability to process and present the vaccine antigen *via* MHC class I molecules, to induce the priming of CTL responses. Soluble protein antigens are usually only inefficiently cross-presented, but particulate antigens that are similar in size to, or larger than, viruses efficiently reach the MHC class I pathway without intracellular replication, and their peptides are efficiently cross-presented (3–4 orders of magnitude higher than peptides from soluble antigens). Processing of particulate antigens that are 20 nm–3 μm in diameter has been documented and there is no clear indication that there are preferred sizes within this range. It is therefore possible that APCs have evolved to effectively process any antigen with dimensions that are similar to pathogens, ranging from viruses (20–100 nm) to bacteria and even cells (in the micrometer range). So, generating particulate vaccines may enhance the uptake, processing and presentation of the antigens by APCs.

Vaccines are usually injected subcutaneously or intramuscularly. However, adaptive immune responses are mainly induced in the secondary lymphoid organs. So, the transport of antigens through the lymphatic system from the peripheral tissues to lymphoid organs is important for vaccine design. Molecules of $20 - 200$ nm efficiently enter the lymphatic system through diffusion and drainage, with an optimal size being

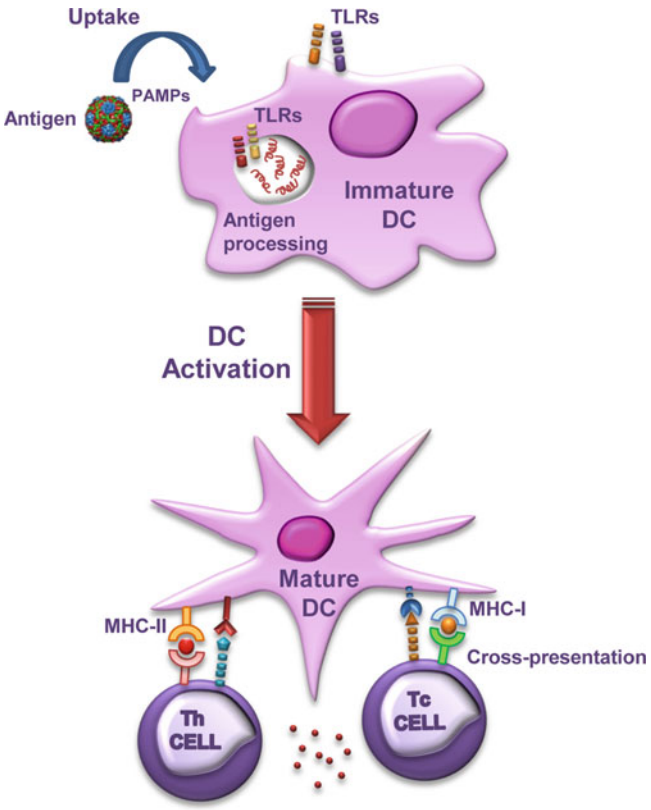


Fig. 21.2 Innate and adaptive immune systems are intimately connected. PAMPs present at the surface of pathogens stimulate antigen uptake and processing by special subsets of APCs, DCs, resulting in upregulation of DC maturation markers. Mature DCs present pathogen-derived antigens to naive Th and Tc cells *via* MHC class-II and class-I molecules, respectively. Secretion of cytokines by DCs stimulates differentiation into B and T effector cells resulting in antibody release and CTL responses

~40 nm, just within the size range of viruses and VLPs. By contrast, particles that are larger than 200 – 500 nm do not efficiently enter lymph capillaries in a free form. Instead, large particles need to be carried into the lymphatic system by specialised cells (particular subsets of APCs), which can squeeze through openings between overlapping endothelial cells. Therefore, size determines whether particles freely drain to lymphoid organs or arrive at these sites in association with cells, and this has major implications for vaccine design and immune cell targeting. As antigens with virus-like dimensions (and smaller) most efficiently reach the lymph nodes in a cell-free state, vaccine delivery systems with similar dimensions will facilitate direct interaction of antigens, in an authentic native configuration, with follicular B cells. This property dramatically increases their ability to induce antibody responses, and it is crucial for the induction of antibodies recognizing conformation-dependent epitopes. In addition, for the induction

of CTL responses, specific subsets of DCs ($CD8^+$ lymphoid DCs) that are essentially restricted to lymphoid organs, must be loaded with antigen. Hence, again particulate antigens with virus-like dimensions are preferable, since antigen size limits the ability of antigens to directly reach the adequate lymph node compartments.

Regarding the structure of antigens, most viral surfaces, as well as bacterial components such as the flagellum, consist of one or a few proteins and consequently are repetitive and highly organised in nature. Both the innate and adaptive immune systems recognise such highly repetitive structures as PAMPs. B cells, for example, recognise specifically and respond strongly to the ordered structure characteristic of viruses [30]. Most importantly, highly repetitive surface patterns efficiently cross-link B cell receptors on the surface of B cells, which represents a strong activation signal for these cells, leading to a strong antibody response [31]. It has been shown that 15–20 immunogenic molecules (haptenated polymer molecules) that are spaced 5–10 nm apart, is an ideal geometry for optimal B cell activation and this is similar to the average spacing of viral coat proteins in virion surfaces [32]. This also holds true for peptides that are conjugated to VLPs, where similar numbers and distances apply [31, 32]. Moreover, the stimulation of B cells by repetitive, highly organised antigens can be strong enough to elicit Th cell-independent induction of antibody responses (T-cell independent B-cell antigens). Therefore, these antigens can activate B cells more efficiently and at much lower concentrations than monomeric antigens.

Finally, it has become apparent that the innate and adaptive immune systems are intimately connected (Fig. 21.2). No relevant adaptive immune response will be mounted unless certain types of APCs, like DCs, are activated, resulting in a rapid release of inflammatory cytokines, and up-regulation of costimulatory molecules, which all serve to condition and shape the subsequent adaptive response [33]. For instance, cross-presentation alone is not sufficient to drive induction of a CTL response, as DCs need to be previously activated (indeed, non-activated DCs presenting antigen to T cells induce T cell tolerance rather than immunity). In this regard, several viruses and VLPs have been shown to induce the phenotypic and functional maturation (activation) of DCs [34], by triggering TLRs which recognise viral-specific PAMPs (Fig. 21.2). Interestingly, human papillomavirus (HPV) derived-VLPs, but not unassembled recombinant L1 protein (the constituent monomer of HPV VLPs), can activate DCs directly [35], suggesting that the size and organised structure of VLPs are important parameters to induce DC activation. However, it should be noted that different VLPs diverge in their ability to induce maturation of DCs, even if the VLPs are similar in size and structure (*i.e.*, VLPs from different polyomavirus species) [36], indicating that other factors that are VLP-specific can also affect the outcome of the interaction between VLPs and DCs.

Alternatively, an interesting approach currently being investigated is the use of PAMPs, in particular TLR ligands, as immune modulating adjuvants. These molecules have the ability to enhance B and T-cell responses induced by an antigen, offering advantages over existing conventional vaccine adjuvants. Virus-related PAMPs being developed as immune modulating adjuvants include CpG-containing oligodeoxynucleotides (CpG-ODNs), double stranded RNA derivatives and

analogues of poly(I:C). Unfortunately, the use of such TLR ligands as vaccine adjuvants is associated with several problematic issues. Because TLR ligands activate their cognate receptors which are present in many types of immune and non-immune cells, they have the potential to be toxic. Thus, co-administration of TLR ligands and antigen by simple mixing may lead to a degree of systemic exposure causing non-specific stimulation of the immune system, with resultant unwanted reactogenicity or immune related pathology. A strategy to solve these problems is to target the TLR ligands directly to DCs. This can be achieved by packaging viral TLR ligands within a particulate delivery system, providing an environment in which the ligand is stable and importantly, preventing its systemic distribution. Moreover, when combined with the antigen, the particulate nature of the delivery system serves to target both the antigen and the TLR ligand to the APCs, conveying them to the appropriate intracellular compartments, thus eliciting highly efficient activation of innate and adaptive immunity. VLPs, virosomes and liposomes have been used to co-deliver antigens with a variety of TLR ligands [22, 23, 31].

In summary, to induce potent immune responses, vaccines of viral size are preferable, as only particles in the nanometer range can reach lymph nodes and directly interact with B cells. Moreover, the antigenic epitopes should be displayed on the nanoparticles in an ordered and highly repetitive way to optimally activate B cells. Nanoparticle-sized vaccines are also preferable for the induction of T cell responses, as size limits the ability of particles to reach the appropriate DC subsets in lymphoid organs. Viruses (and some VLPs) have the ability to directly induce the activation of DCs by triggering TLRs. Delivery of TLR ligands by nanoparticles together with the antigen may facilitate activation of DCs, which is essential for the induction of T cell responses. Hence, vaccine carrier systems, such as VLPs, that mimic the size, geometry and PAMPs of viruses may be one possible way to optimally harness viral immunogenic properties, without the risks associated with viral infection (Fig. 21.3).

21.4 Virus-Like Particles (VLPs) as Antiviral Vaccines

21.4.1 Characteristics of VLPs

VLPs are supramolecular assemblies with a well-defined geometry, usually icosahedrons or rod-like structures, and diameters in the range of 20–120 nm [37] (Fig. 21.4). They are based on the natural intrinsic ability of many viral structural proteins to spontaneously assemble into multimeric structures when expressed using recombinant expression systems. VLPs mimic the organization and conformation of authentic native viruses (Fig. 21.5), with or without lipid envelope depending on the virus of origin and the viral structural protein expressed (*i.e.*, capsid proteins or viral envelope proteins), or in some cases subviral particles. Thus, VLPs are usually

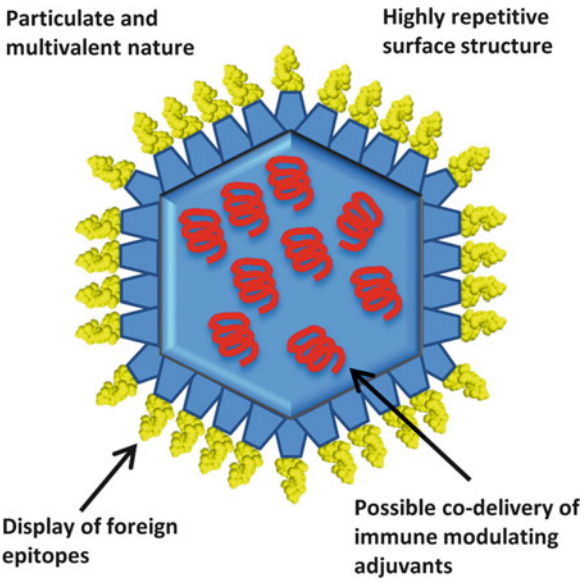


Fig. 21.3 Immunogenic features of VLPs presenting foreign antigens

structurally and antigenically indistinguishable from infectious virus particles, generally retaining the ability to bind and penetrate host cells, and are consequently highly immunogenic, but unable to replicate, since they lack the viral genome. In addition, immunization with VLP-based vaccines does not induce antibody responses to internal or non-structural viral proteins (which do not form part of the VLPs but are present in virus-infected cells, and therefore induce antibody responses in the infected host). This enables serological differentiation between vaccinated and virus-infected animals (marker vaccines), which represents an important feature for vaccines intended for veterinary use against notifiable diseases of livestock. Overall, these advantages have made VLPs attractive vaccine candidates against many viral diseases [8–13, 38–40].

The generation of VLPs derived from certain viruses has been strongly motivated by the lack of efficient cell culture propagation systems for those viruses, as was the case for hepatitis B virus (HBV), hepatitis C virus (HCV), HPV, or caliciviruses. An additional important reason for these efforts has been the possibility to substitute infectious viruses requiring high-level biosafety containment laboratories for their handling, like Ebola virus or Severe Acute Respiratory Syndrome virus, by VLPs. Hence, those VLPs have been used as efficient surrogates for virus diagnostics, basic research applications (like understanding the assembly or architecture of these viruses), and vaccine development. Moreover, an important challenge nowadays consists in the substitution of live vaccines, for example against influenza A virus, by an appropriate highly efficient and safe VLP-based vaccine.

In some instances, the VLPs are derived from internal structural viral proteins, like the HBV core antigen (HBcAg), or the Gag capsid proteins from retroviruses

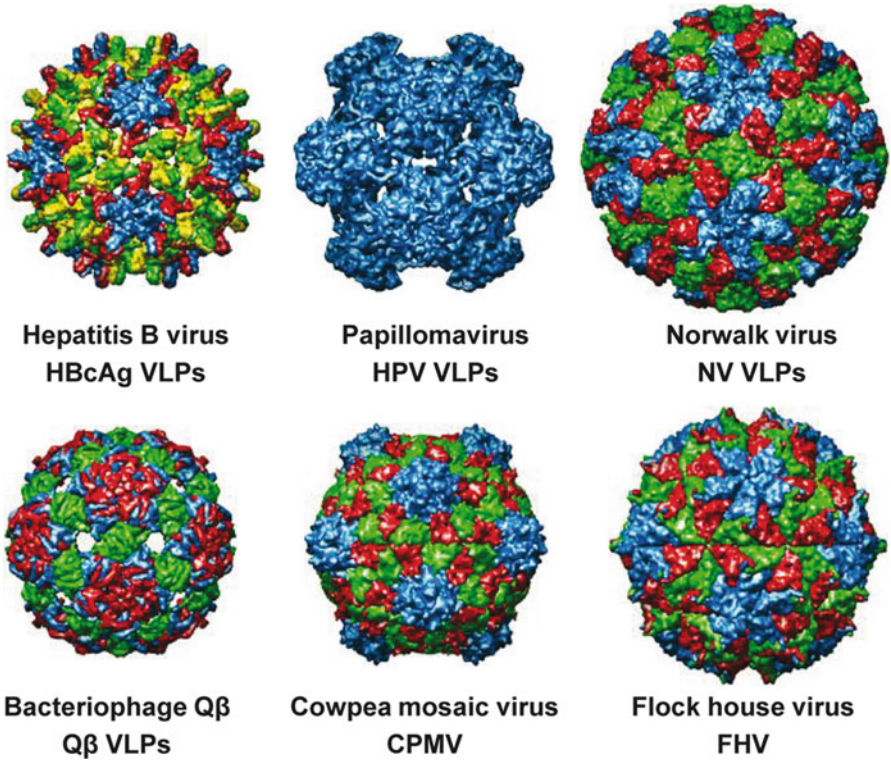


Fig. 21.4 Examples of viruses and VLPs mentioned in this chapter that have been developed as VLP-based vaccines or platforms for foreign antigen display. The images are coloured to distinguish the symmetry-related subunits and are taken from the VIPER database (http://viperdb.scripps.edu). Protein Data Bank (PDB) IDs: hepatitis B capsid, *1QGT*; Papillomavirus capsid, *1DZL*; Norwalk virus capsid, *1IHM*; bacteriophage Qβ, *1QBE*; cowpea mosaic virus, *1NY7*; flock house virus, 2Z2Q

(including HIV), which are capable of self-assembling into subviral particles. Such VLPs do not mimic the external surface structure of the corresponding native virions, and antibodies induced by them are not able to interfere with the course of virus infection. Hence, these VLPs are not suited *per se* for vaccine development against the viruses from which they were derived, but they can be used as scaffolds for foreign antigen presentation (see Sect. 21.5). Of the utmost interest for the development of antiviral vaccines are VLPs composed of viral surface proteins, usually corresponding to the external capsids of nonenveloped virions, particularly when virus-neutralising antibodies recognise those proteins and/or their assemblies. Examples of such VLPs are those derived from HPV, bluetongue virus (BTV), rotavirus, parvovirus, calicivirus, and infectious bursal disease virus (IBDV) [9, 13, 15, 17].

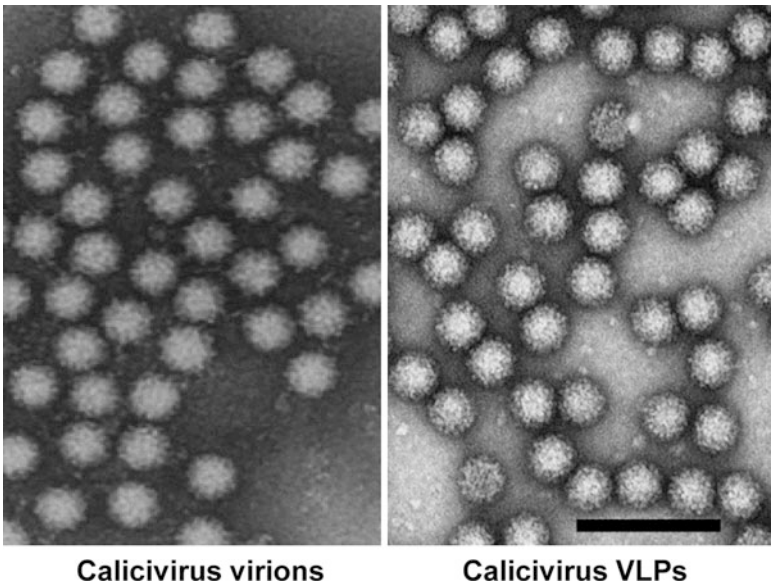


Fig. 21.5 Electron microscopy of negatively stained calicivirus particles: virions (*left*) and VLPs (*right*). Scale bar = 100 nm

21.4.2 Structural Classification of VLPs

VLPs have been produced for a wide range of taxonomically and structurally distinct viruses that infect humans and other animals, as well as plant and yeast viruses, and bacteriophages [13, 17, 41, 42]. These comprise viruses that have a single capsid protein, multiple capsid proteins, and those with or without lipid envelopes, indicating that the ability to develop VLPs does not appear to be limited to any type of virus family or by the complexity or structure of the virus particle. The largest number of VLPs have been obtained from viruses with single-stranded RNA genomes of positive polarity. Numerous families and genera of this most diverse structural class of viruses have been involved in the construction of differently sized and structured VLPs. Regarding organization and symmetry (see Chap. 2), this group of VLPs is represented by structures of two different symmetries: icosahedral (prevalent number of prototypes, 18 in total) and rod-like (three prototypes), derived from nonenveloped viruses, as well as by isometric VLPs without any distinct symmetry (three prototypes) derived from enveloped viruses [42].

Structural variants of VLPs are depicted in (Fig. 21.6). The simplest variant is represented by the self-assembly of a single viral structural protein into a single-shelled VLP. One of the most studied VLPs of structurally simple viruses is the HPV-derived VLP. HPV virions contain the major and minor capsid proteins (L1 and L2, respectively). The virus capsid (60 nm) contains 72 pentamers of L1

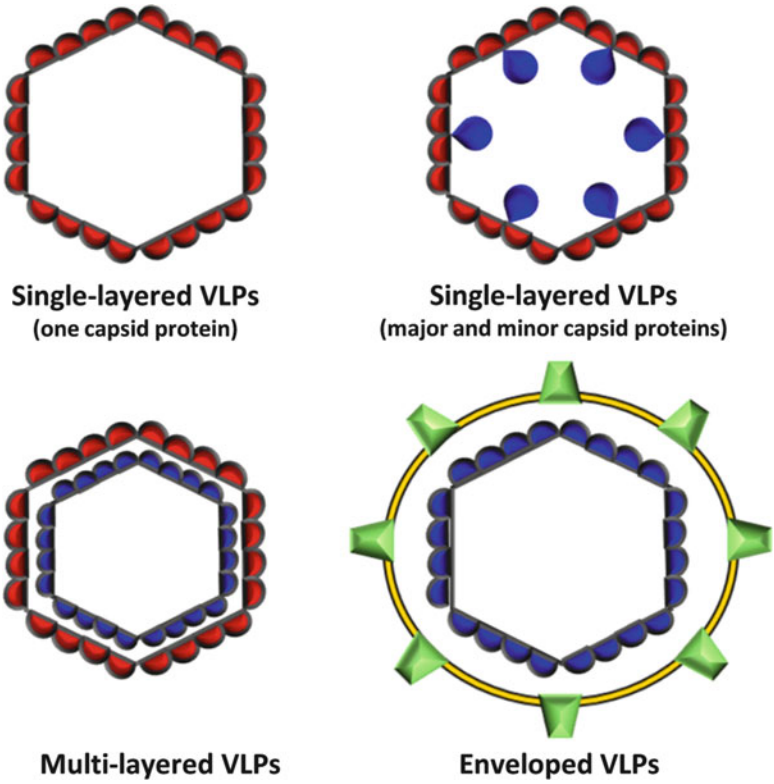


Fig. 21.6 Structural variants of VLPs

protein, centered on the vertices of a T = 7 icosahedral lattice (see Chap. 2), while L2 protein is present at about 1/30 of the abundance of L1. HPV VLPs are usually formed just by L1 protein, with morphological characteristics very similar to those of HPV virus capsids [43]. Parvovirus VLPs are also formed by a single protein (60 copies of VP2, the major viral structural protein). Other examples of VLPs formed by a single viral structural protein are those derived from the HBcAg, hepatitis E virus (HEV), calicivirus, nodavirus, circovirus, or bacteriophage Qβ [9, 17, 42]. Similarly, VLPs can be composed by both, a major and a minor protein forming a single shell, like in the case of HPV-derived VLPs composed by L1 and L2 proteins [44].

More complex multi-layered VLPs are generated by the co-expression of several viral structural proteins, like in the case of the BTV-derived VLPs (*Reoviridae* family). The icosahedral BTV virions are formed from non-equimolar amounts of seven structural proteins (VP1-VP7). However, only the four major structural proteins of the virus (VP2, VP3, VP5 and VP7) are necessary for the formation of double-layered VLPs, which are structurally indistinguishable from authentic virus particles [13]. VLPs derived from rotaviruses, also belonging to the *Reoviridae*

family, can be generated by co-expressing different subsets of viral structural proteins [11]. Expressing VP2 alone produces single-layered VLPs. Co-expression of proteins VP2 and VP6 that form the innermost and middle capsid layers of rotavirus particles, respectively, produces double-layered 2/6-VLPs. Co-expression of VP4 with VP2 and VP6 yields double-layered VLPs with VP4 spikes (2/4/6-VLPs). Co-expression of VP2, VP6, and VP7 produces triple-layered 2/6/7-VLPs, and finally, co-expression of VP2, VP4, VP6, and VP7 results in 2/4/6/7-VLPs that are complete triple-layered VLPs with the VP4 spike protein.

Another prototypic example of complex VLPs are enveloped VLPs containing unilamellar liposomes with embedded viral envelope proteins. The first VLP vaccine to be produced and characterised (in the early 1980s) consisted of isometric, spherical, pleomorphic particles with diameter between 17 and 25 nm, formed by the HBV small surface antigen (HBsAg), co-assembled with host cellular membranes [42]. Several other enveloped VLPs have been successfully developed by the simultaneous synthesis and co-assembly of two different viral protein species, *i.e.*, capsid and envelope proteins. This is the approach used for the generation of VLPs derived from influenza virus, which contain both viral glycoproteins: hemagglutinin and neuraminidase, as well as matrix structural proteins. Other examples of enveloped VLPs are those derived from HCV, formed by the co-expression of core, E1 and E2 proteins, and retrovirus-based VLPs (like those of HIV), containing the Gag capsid protein and the envelope protein (gp120). The VLPs derived from viruses with lipid envelopes are sometimes referred to as virosomes [45].

21.4.3 Expression Systems for VLP Production

VLPs can be produced in many different production platforms, including mammalian cells, insect cells, plants, yeast, bacteria, and recombinant expression systems based on viral vectors: baculovirus, poxviruses, alphavirus replicons, etc. [41]. Currently, the most popular choice is expression in insect cells using the recombinant baculovirus technology [9, 46, 47]. This expression system has many advantages for VLP production. Large amounts of correctly folded recombinant proteins can be produced with mammalian-like post-translational modifications. Although yeast and bacteria cells can achieve similar yields, the complexity of the VLPs produced with the baculovirus expression system is remarkably higher (VLPs formed from the co-expression of up to five proteins). An additional advantage is that baculoviruses have a limited host range (namely for insects) and are hence safe for vertebrates. The design of recombinant baculoviruses is simple and fast, providing a high versatility to this expression system. This is important when producing vaccines for viruses whose surface proteins rapidly mutate, (*i.e.*, influenza A virus), a fundamental requirement to contend with potential pandemics in a timely manner.

Although VLPs have been produced for a wide range of viruses, as indicated above, clearly not all are equally suitable for the development of vaccines. Even if a VLP-based product candidate has been shown to elicit a robust immune response under laboratory conditions, it might not be developed as a vaccine for widespread use, if its manufacturing process is not scalable or cost-effective. According to the experiences accumulated in the past years, some predictive considerations about the chances of success in adequate VLP production yields can be enumerated [41, 48]. Structurally simple non-enveloped icosahedral viruses are the most suitable candidates for successful high yield production of VLPs. VLPs made by the assembly of a single protein are usually able to be produced in large amount and high quality (*i.e.* HBcAg, parvovirus, calicivirus or papillomavirus, bacteriophage Qβ). An increase in the number of proteins to be expressed and assembled proportionally raises the difficulties of correct assembly [41], usually leading to lower yields of VLP production. The efficient assembly of capsids derived from proteins encoded by multiple discrete mRNAs (*i.e.*, orbivirus, birnavirus), requires the co-expression of the interacting proteins in the vicinity of each other and within the same cell. On the other hand, capsid proteins that need to be processed from a polyprotein precursor in a timely fashion (monocystronic RNAs, *i.e.*, picornavirus, flavivirus) are usually assembled with low yields, probably due to relatively inefficient processing of the precursors in recombinant expression systems and/or the toxicity of the viral protease required for processing. Finally, due to the inherent properties of the lipid envelope, production of enveloped VLPs is technically complex [41]. However, despite these considerations, progresses are being made, and it is expected that in the near future the integration of process optimization tools (*i.e.*, molecular biology, genetic engineering and systems biology), will allow to overcome some of the current limitations affecting the large scale production of several types of VLPs [46, 47, 49, 50].

21.4.4 VLPs as Stand-Alone Vaccines

The most straightforward application of VLPs is to use them for vaccination against the virus from which they were derived. Currently, VLP-based vaccines against human diseases are in various stages of development, spanning from preclinical evaluation to market (Table 21.1) [9, 16, 17, 24].

VLP-based vaccines for HBV have been licensed commercially. Recombivax® (Merck) and Engerix® (GlaxoSmithKline), both licensed in 1986, are subvirion 22-nm spherical lipid-protein particles consisting of self-assemblages of the HBV small surface protein, HBsAg, co-assembled with membrane lipids, expressed in genetically engineered yeast (*Saccharomyces cerevisiae*). Although these vaccines were proved to be highly effective, they suffered from a lack of immunogenicity in a small percentage of patients (~5–10 % nonresponders), which was determined to be due to an absence of epitopes derived from PreS proteins (PreS1 and PreS2) on the surface of the VLPs. A more immunogenic VLP vaccine was subsequently

Table 21.1 Selected VLPs tested as vaccine candidates for different human and animal diseases

Virus	Family	Recombinant proteins	Envelope	Morphology of VLPs	Target species
Hepatitis B virus	*Hepadnaviridae*	Surface antigen	Yes	Isometric	Human
Human papillomavirus	*Papillomaviridae*	L1	No	Icosahedral, T = 7d, 1	Human
Influenza A	*Orthomyxoviridae*	HA, NA, M1, M2	Yes	Pleomorphic	Human
Norwalk virus	*Caliciviridae*	Capsid protein	No	Icosahedral, T = 3	Human
Hepatitis C virus	*Flaviviridae*	Core, E1, E2	Yes	Isometric	Human
Rotavirus	*Reoviridae*	VP2, VP4, VP6, VP7	No	Icosahedral, T = 13	Human
Rift valley fever virus	*Bunyaviridae*	N, Gn, Gc	Yes	Pleomorphic	Human
Porcine circovirus type 2	*Circoviridae*	Coat protein	No	Icosahedral, T = 1	Pig
Bluetongue virus	*Reoviridae*	VP2, VP3, VP5, VP7	No	Icosahedral, T = 13 l	Sheep
Rabbit hemorrhagic disease virus	*Caliciviridae*	Capsid protein	No	Icosahedral, T = 3	Rabbit
Porcine parvovirus, canine parvovirus, mink enteritis virus	*Parvoviridae*	VP2	No	Icosahedral, T = 1	Pig, dog, mink
Infectious bursal disease virus	*Birnaviridae*	VP2, VP3, VP4	No	Icosahedral, T = 13, 1	Chicken
Newcastle disease virus	*Paramyxoviridae*	NP, M, F, HN	Yes	Pleomorphic	Chicken
Avian influenza virus	*Orthomyxoviridae*	HA, NA, M1	Yes	Pleomorphic	Chicken
Nodavirus	*Nodaviridae*	Coat protein	No	Icosahedral, T = 3	Fish

developed that contained, in addition to HBsAg, the proteins PreS1 and PreS2, which are present in the native HBV envelope at lower levels than HBsAg. This new HBV VLP-based vaccine (Sci-B-Vac®, SciGene), expressed in the mammalian CHO cell line, was found to elicit a strong antibody response and 100 % seroconversion and seroprotection rates. Hence, a better mimic of the authentic virion surface structure resulted in broadened effectiveness of the HBV VLP-based vaccine [42].

The second type of commercially available VLP-based vaccines was developed against HPV, for the prevention of cervical cancer. These vaccines consist of 40 nm VLPs assembled from the HPV major capsid protein, L1. Gardasil® (Merck) is produced in yeast (*Saccharomyces cerevisiae*) and was licensed in 2006. Cervarix® (GlaxoSmithKline), is produced using the recombinant baculovirus expression

system and was licensed in 2009. Both contain mixtures of recombinant VLPs derived from different HPV types associated with genital infection (HPV types 6, 11, 16 and 18). The vaccines show great promise, as are almost completely protective against HPV types from which they are derived, and achieve partial protection against other phylogenetically related types. Since HPV types 16 and 18 are implicated in 70 % of cervical cancers, these vaccines are considered a breakthrough for preventing cervical cancer and are expected to drastically reduce the occurrence of this life-threatening disease in women [44].

Other VLP-based candidate vaccines for human diseases in an advanced stage of clinical development include those directed against human parvovirus (B19), Norwalk virus, rotavirus and influenza A virus [9, 16, 17, 24]. In the veterinary field there is currently one VLP-based vaccine commercially available against porcine circovirus type 2 (PCV2), Porcilis PCV® (Intervet), which was licensed in 2009. Other promising candidate vaccines in different stages of development include those against: BTV, IBDV, rabbit hemorrhagic disease virus, canine and porcine parvovirus, nodavirus, Rift valley fever virus, avian influenza virus or Newcastle disease virus [12, 38].

21.5 VLPs as Platforms for Foreign Antigen Display. Structure-Based Engineering of VLPs for Vaccine Development

As indicated in previous sections, VLPs can also be used as platforms for the multimeric display of foreign antigens, that can be incorporated into VLPs either by genetic fusion (chimeric VLPs) or by chemical conjugation (conjugated VLPs). In such cases VLPs serve both, as a presentation scaffold for antigens derived from other pathogens, in a suitable repetitive and highly organised configuration, and as adjuvants to boost the immune response. Ideally, the underlying immunogenic ‘viral fingerprint’ of the VLP is imparted to the attached antigen, making it as potent an immunogen as the VLP itself.

The poor immune response of many soluble antigens can be overcome by rendering them highly repetitive incorporating them to a carrier protein [51]. In contrast to monomeric and oligomeric carrier proteins commonly used for antigen presentation, such as bovine serum albumin and keyhole limpet hemocyanin, chimeric VLP structures are able to provide not only a high density of introduced foreign antigens per particle, but also a distinctive three-dimensional conformation, which is especially important for the presentation of conformational epitopes. Hence, the regular, repetitive pattern and correct conformation of inserted epitopes, are factors encouraging the development of VLPs as platforms for inducing immunological response against foreign antigens.

Insertion in specific sites of VLPs of continuous epitopes (those included in a short peptide segment) or continuous parts of discontinuous epitopes (those made up of several peptide segments in defined conformations) may alter the

conformational preferences of the peptide segment relative to the native ones (*i.e.*, in their native protein or viral capsid), which may have dramatic effects on their immunogenicity as a part of the chimeric VLP. If complete discontinuous epitopes must be inserted in the VLP, the problem is severely compounded, as different peptide segments must be inserted in such a way that not only their conformational preferences, but also their geometric relationships are preserved in the chimeric VLP, relative to those in the native protein or viral particle. Detailed structural information and computational docking approaches may be critical for the rational design of highly immunogenic chimeric VLPs.

VLP chimeras have been extensively explored as vaccine candidates since mid-1980s [42, 52]. The interest in rational-based manipulations on chimeric VLPs was reinforced by the simultaneous development of recombinant DNA engineering techniques and structural knowledge about virions, capsids, and envelope proteins obtained by the use of high-resolution techniques (X-ray crystallography (see Chap. 4) and cryo-electron microscopy (cryo-EM) (see Chap. 3) [53]. Concurrently, research in immunology, molecular biology and structural studies led to the identification, characterisation and fine mapping of immunogenic epitopes derived from pathogens causing relevant diseases. X-ray crystallographic studies of antigen-antibody complexes as well as peptide-T-cell receptor interactions provided considerable insight into the structural basis of immunological recognition [54].

Currently a wealth of high-resolution viral structural information (see Chaps. 2, 3, 4, 5, 6 and 7) has facilitated the ability to modify VLPs deliberately so that they function essentially as molecular scaffolds for antigen presentation. VLPs derived from both double-and-single-stranded DNA and RNA viruses encompassing 14 different families of virus have been successfully used for the display of foreign antigens [14, 15, 17, 42] (Table 21.2). Results obtained using VLPs derived from different viruses, like HBV, HPV, parvovirus, calicivirus or bacteriophages have illustrated the potential of this approach to induce strong immune responses against foreign B and T-cell epitopes [14, 16, 17].

21.5.1 Chimeric VLPs Inducing Antibody Responses Against Target Molecules

In order to induce high-titer antibody responses effectively, target antigens must be displayed on the surface of VLPs, in immunodominant regions, at a high density. The genetic insertion of target sequences into viral structural proteins to generate chimeric particles is the most common method for displaying foreign epitopes on VLPs. Essentially, DNA fragments encoding continuous or discontinuous immunological epitopes are cloned into the genes encoding the self-assembly-competent polypeptides that form the VLPs. Upon assembly of the chimeric subunit proteins into supramolecular structures, the introduced epitopes are presented at a relatively high density (usually one copy per VLP subunit) and, ideally, in an accessible and

Table 21.2 Examples of VLPs and chimeric virions used as platforms for foreign antigen display

VLP platform or Viral nanoparticle	Foreign antigen	Incorporation method
HBcAg	Influenza A M2 extracellular domain, B- and T-cell epitopes of HCV, epitopes derived from malaria	Genetic fusion
HBsAg	Dengue virus envelope, HIV gp41epitopes,	Genetic fusion
Bovine papillomavirus	CTL epitopes of HPV and HIV, L2 HPV epitopes, Aβ peptide	Genetic fusion
HPV	HPV E7 oncoprotein	Genetic fusion
HPV	Aβ peptide	Chemical conjugation
Rabbit hemorrhagic disease virus	Model T-cell epitopes (*i.e.*, Chicken ovoalbumin CTL epitope)	Genetic fusion and chemical conjugation
Porcine parvovirus	Model T-cell epitopes (*i.e.*, CTL epitope derived from LCMV nucleoprotein)	Genetic fusion
Phage Qβ	Nicotine, angiotensin II, CCR5 extracellular domain	Chemical conjugation
CPMV	B-cell epitope (NIm-1A) of HRV-14, B-cell epitope of canine parvovirus	Chemical conjugation
FHV	ANTXR2 domain, ectodomain of G protein from Rift valley fever virus	Genetic fusion

conformationally relevant manner. An advantage of this chimeric VLP technique is that successful incorporation of a target epitope guarantees that the antigen will be displayed in the same conformation and at high density on the particle surface. This technique also has substantial advantages from a manufacturing standpoint. Chimeric particles can be purified using the same well-established methods used to purify unmodified parental VLPs.

The key to this technology is the identification of adequate permissive insertion sites to incorporate the target epitopes, within the primary sequence of the VLP subunits. That is, sites that do not compromise the correct folding and assembly of the VLP protein subunits, and do not alter the structural integrity and immunogenicity of the VLP. Obviously, depending on the type of foreign epitope to be displayed, the structural requirements of the insertion sites would be different. B-cell epitopes should be located at exposed sites on the surface of the particles, preferentially located at immunodominant regions, readily accessible for direct interaction with B-cells. However, T-cell epitopes do not need to be located at exposed sites within the VLP structure, because these epitopes have to be proteolytically processed by APCs before being presented to the target T-cells (as explained in Sect. 21.3), and thus can be inserted at nonaccessible permissive sites of the VLP. Hence, knowledge of the three-dimensional structure of the VLP scaffold is a critical requirement for successful generation of chimeric VLPs. If the three-dimensional structure of the VLP is not available, attempts can be made to locate suitable insertion sites in surface regions predicted on the basis of sequence analysis, generally at the N or C termini of the VLP protein subunits [55]. These sites are expected to raise fewer assembly problems for the resulting chimeric

VLPs. However, the immune responses obtained against the inserted foreign epitopes are often limited, most likely because the epitopes are not optimally exposed to the immune system at that locations.

Since many B-cell epitopes in their native environments are found at surface loops of the pathogen's proteins, this is often a good choice for the insertion of foreign antigens in VLP protein subunits. In some cases, proper folding and presentation can be accomplished by replacing exposed, immunodominant viral epitopes located at loop structures on VLPs, with the desired target epitope. However, generating chimeric VLPs is still largely empirical; it is almost impossible to predict whether individual peptides will be compatible with VLP assembly at defined insertion sites or whether the insertions will be immunogenic in the resulting chimeric VLP. For instance, it has been shown that minor displacements in the insertion site (two or three amino acid positions) can cause drastic changes in the immunogenicity of the inserted sequence [56]. Peptide insertions with high hydrophobicity, a high β-strand index, a large volume or a strong positive charge are prone to cause problems in VLP assembly. The success rate can be improved using combinatorial technologies. This type of approach was used to generate chimeric VLPs derived from the core protein of woodchuck hepadnavirus [57], enabling the identification of 17 different sites in the subunit protein that could be used to insert target sequences.

Another limitation attributed to the chimeric approach is that the size of the antigens that can be inserted into VLPs, in particular into their surface-exposed immunodominant regions, is usually restricted to short peptides. Chimeric protein subunits containing foreign peptide insertions longer than 20–30 amino acids often fail to assemble into VLPs. However, it should be noted that there are marked differences in the versatility of the different VLPs as antigen display platforms, regarding the size of the insertions they can tolerate. Relatively large insertions have been successfully incorporated into exposed sites in VLPs, like the 238 amino acids of the complete green fluorescent protein (GFP), that has been inserted in HBcAg-derived VLPs [42]. Other whole proteins or complete protein domains, with more relevance as immunogens to induce antibody responses, have been successfully inserted in chimeric VLPs, like a 120 amino acid-long polypeptide from the hantavirus nucleocapsid protein in HBcAg-derived VLPs, or a 395 amino acid portion of dengue virus envelope protein type 2, in HBsAg-derived VLPs [17]. Hence, in some cases, it is possible to display large antigens in the surface of chimeric VLPs. This might be important in order to increase the number of epitopes targeted by an individual chimeric VLP, or for targeting complex discontinuous B-cell epitopes.

The VLP derived from the HBcAg is the most extensively investigated chimeric carrier of foreign antigens to date [42, 52]. HBcAg polypeptide is 183–185 amino acids long (21-kDa). The organization of HBcAg derived particles is largely α-helical and thus, quite different from previously known viral capsid proteins with β-sheet jellyroll packings. The HBcAg monomer fold is stabilised by a hydrophobic core that is highly conserved among human HBV variants. Association of two amphipathic α-helical hairpins results in the formation of a dimer with a

four-helix bundle as the major central feature. The dimers are able to assemble into two types of particles, large and small, that are 34 and 30 nm in diameter and correspond to triangulation numbers T = 4 and T = 3, containing 240 and 180 monomers, respectively (see Chap. 2). The major immunodominant region (MIR) with the central amino acid positions 76–81 is located at the tips of the α-helical hairpins that form spikes on the capsid surface. In addition to the highly exposed MIR, the region between amino acids 127 and 133 is the next exposed and accessible epitope on the particle surface. This region is located at the end of the C-terminal α-helix and forms small protrusions on the surface of the HBcAg particle. Foreign antigens can be inserted at the N- and C-termini, and at the MIR (*i.e.*, the tips of the VLP spikes), without affecting the ability of the resulting chimeric proteins to assemble into VLPs [42, 52]. The N-terminal end of HBcAg has been widely used as insertion site to display a variety of short foreign epitopes. The capacity for incorporating foreign peptides at this site is around 50 amino acids. The inserted epitopes are accessible to specific antibodies and in general terms, high levels of specific antibody responses are achieved. Regarding the C-terminal insertions, amino acid positions 144, 149, 156, 163 and 167 are the most frequently used sites for the incorporation of foreign epitopes. The capacity of the C-terminal end to accept foreign insertions usually exceeds 100 amino acids residues, depending on the structure of the inserted polypeptide. In some cases, quite large polypeptides have been successfully inserted at this location. A 559 amino acid long insertion, containing three copies of the HCV core protein, did not prevent self-assembly of chimeric proteins. However, the C-terminal end of HBcAg is not exposed at the surface of the VLPs and thus, chimeric constructions harbouring foreign antigens at this site usually induce weak to moderate specific antibody responses. Hence, this site is mainly used for the insertion of T-cell epitopes. The most interesting and promising site for foreign insertions is the MIR, which allows full exposure of the inserted peptides on the VLP surface. In addition, this insertion site has revealed an unexpected high capacity for the incorporation of large foreign polypeptides. The entire 120 amino acids long immunoprotective region of the hantavirus nucleocapsid protein, and the 238 amino acids long GFP (as indicated above), have been successfully incorporated into chimeric HBcAg VLPs at this insertion site.

Mosaic VLPs have been developed to aid in the assembly of complex VLP chimeras. These VLPs are typically formed by co-expression of a wild-type VLP subunit along with a chimeric VLP subunit. The resultant VLPs contain copies of both wild-type and chimeric subunits which are incorporated with differing proportions. The use of mosaic VLPs has enabled the formation of HBcAg VLPs containing fragments of up to 213 amino acids from the hantavirus nucleoprotein, or 163 amino acids from the HBV preS domain [42]. Alternatively, coexpression of VLP subunit proteins derived from different virus strains may be used to assemble mosaic VLP vaccines that generate antibodies that protect against multiple pathogen serotypes. This approach has been used for the generation of mosaic VLPs derived from HPV formed by the co-expression of L1 and L2 capsid proteins of both, HPV-6 and HPV-16 types. The resulting VLPs comprised all four co-assembled subunit proteins.

21.5.2 Chimeric VLPs Targeting Self-Antigens

Traditionally, antibody-inducing vaccines were used solely to provide prophylactic protection against pathogens. More recently, there has been a growing interest in the development of new vaccines with the goal of deliberately inducing antibody responses against self-molecules that are involved in chronic disease processes. These autoantibodies are expected to serve as competitive inhibitors to fight diseases. The ability to induce antibody responses against self-molecules is seemingly limited by the mechanisms of B-cell tolerance, which eliminate, desensitise or change the specificity of potentially self-reactive B cells, in order to prevent autoimmune disorders. However, B-cell tolerance is somehow inefficient; it has been estimated that approximately 20 % of long-lived mature B cells are self-reactive. Thus, antibodies against self-molecules can be induced by vaccination.

The main hurdle for eliciting a high titer antibody response against self-antigens is overcoming T-cell tolerance. Stringent T-cell tolerance mechanisms ensure that autoreactive B cells cannot receive T-cell help and, therefore, cannot proliferate and produce potentially damaging autoantibodies. Thus, one strategy for inducing antiself antibodies has been to immunise with self-antigens chemically conjugated or genetically linked to foreign T-helper epitopes. This approach requires large doses of antigen and the use of powerful adjuvants. In addition, antibody titers induced against the self-antigen are often very low and rapidly diminish, particularly in comparison with the responses induced by vaccines against foreign antigens. By contrast, self-antigens displayed on VLPs are inherently immunogenic at low doses and without exogenous adjuvants. It has been shown that self-antigens incorporated to VLPs can induce antibody titers up to 1,000-times higher than those induced by the same self-antigen linked to a foreign T-helper epitope. Moreover, VLP display can make a self-antigen as immunogenic as a foreign antigen presented in the same context. The magnitude of the antiself antibody responses correlates with the density at which the self-antigen is displayed on the VLP surface. In conclusion, B cell unresponsiveness can be overcome by antigens presented in a repetitive and highly organised ‘foreign-like’ manner. In fact, the immune system is largely unable to distinguish between self and foreign proteins based on antigenic epitopes, but does so based on antigenic structural organisation.

VLP display of self-antigens has been successfully used to target molecules that are involved in the pathogenesis of a variety of chronic diseases, including Alzheimer’s disease, rheumatoid arthritis, hypertension, acquired immunodeficiency syndrome and certain cancers [18–20, 24]. Some of these vaccines have shown clinical efficacy in animal models and several are currently in clinical trials. However, a vaccine inducing an antibody response against a self-protein raises safety concerns that must be addressed.

Alzheimer’s disease is a neurodegenerative disease that results in accumulation of plaques consisting of aggregated amyloid-β (Aβ) peptide, neurofibrillary tangles, and loss of neurons leading to dementia. One therapeutic strategy has been the use

of vaccination against Aβ to eliminate the accumulation of plaques in the brain in the hope of preserving neurons. The attachment of an Aβ peptide to papillomavirus VLP capsids induced high levels of specific antibodies and inhibited effective assembly of peptides into neurotoxic peptides *in vitro*. Aβ deposits were also reduced after immunization of a mouse model of Alzheimer's. However, clinical trials of another vaccine formulation against Aβ peptide resulted in development of encephalitis in a subset of patients, that was ascribed to T-cell responses associated with the vaccine. Hence, an effective Alzheimer's disease vaccine might be dependent on the consistent induction of high-titer anti-Aβ antibodies in the absence of inflammatory antiself T-cell responses. A VLP-based vaccine targeting a short peptide at the N-terminus of Aβ, which does not contain a T-cell epitope, could be one possible solution. Two different VLP-based Aβ vaccines targeting this epitope induced high levels of anti-Aβ antibodies, and were not associated with unwanted T-cell responses. Thus, the flexibility of the VLP display system allows the tailoring of vaccines, hopefully enabling to conform to both safety and efficacy requirements.

Another example is one potential therapeutic approach to prevent HIV infection, based on the production of autoantibodies against the HIV cellular receptor CCR5. It is an indirect route to vaccination, avoiding the problem of constant mutation of the viral antigens. A peptide from CCR5 was attached to HPV VLPs and when administered in mice, these particles initiated production of autoantibodies, which inhibited binding of the ligand and blocked infection of an indicator cell line expressing CCR5.

Vaccines directed against self-molecules hold promise for treating a variety of chronic diseases by the induction of antibodies targeting endogenous proteins. In contrast to passive vaccination with monoclonal antibodies, active immunisation with VLP-based vaccines displaying self-peptides induce long-lasting antibody responses, potentially providing affordable and convenient therapy for patients.

21.5.3 Chimeric VLPs Inducing Cell-Mediated Immunity

The use of chimeric VLPs as platforms to induce CTL responses against foreign antigens presents different requirements than vaccines aimed at inducing humoral responses. Unlike B-cell epitopes, T-cell epitopes do not need to be exposed on the surface of the particle or presented at high density, are included in single peptide segments (*i.e.*, they are continuous epitopes), and do not depend on their being stabilised in a specific folded conformation before being recognised by the corresponding immune system molecule. This enables diverse approaches for the generation of chimeric VLPs intended for this purpose. Chimeric VLPs can be generated not only by fusing target antigens to viral major capsid proteins (as were the examples cited in the previous sections), but also by making fusions

between target antigens and proteins that are minor structural components of VLPs, and are often not necessary for particle formation. In this case, the target antigens would not be displayed in high numbers per particle, and will usually be located facing the interior of the VLPs. This is not detrimental in the case of T-cell epitopes, and in fact, in some situations can be advantageous. Thus, targeting large polypeptides might be necessary in situations in which the most relevant target CTL epitopes in a protein have not yet been precisely mapped. Large peptides and even full-length proteins can be added to minor structural proteins without interfering with capsid assembly. For instance, the nonstructural HPV encoded oncoprotein, E7, was fused to the HPV minor capsid protein L2. When co-expressed alongside the HPV major capsid protein (L1), L2–E7 was incorporated efficiently into VLPs and because E7 was incorporated on the inside of the VLP, the fusion domain did not interfere with cell binding. Chimeric L1/L2–E7 VLPs induced strong CTL responses that protected mice from challenge with an E7-expressing tumor cell line [58].

As indicated in Sect. 21.3, in order to activate cytotoxic T-cells, DCs must also provide appropriate costimulatory signals that are only upregulated upon DC maturation. Certain VLP types, such as HPV VLPs, can provoke these signals by directly stimulating the phenotypic and functional maturation of DCs, and thus, are highly efficient inducing CTL responses. In addition, the immunogenicity of many VLPs can be improved by incorporating to the particles substances with adjuvant activity, like TLR ligands. Proof of concept was established using bacteriophage Qβ VLPs co-delivering TLR ligands and CTL epitopes to DCs in animal models [24]. It was shown that Qβ-VLPs devoid of any TLR ligands essentially failed to induce a CTL response, while Qβ-VLPs which were internally loaded with CpG-ODNs induced a robust response against the foreign CTL epitope incorporated to the VLPs (the p33 epitope from lymphocytic choriomeningitis virus glycoprotein).

In summary, numerous chimeric VLPs designed to present B and T-cell epitopes in this manner have been tested in preclinical research [9, 14, 16, 17, 19, 20, 24]. The final intended application of such vaccines spans human to veterinary use and prophylactic to therapeutic treatment. Epitopes have been derived from viral, bacterial, eukaryotic parasitic, and disease related self-molecules. Four of these chimeric vaccines have entered clinical trials: an anti-influenza A M2-HBcAg VLP vaccine, an anti-HIV p17/p24:Ty VLP (a chimeric VLP derived from the yeast retrotransposon Ty1), and two chimeric anti-malaria vaccines based on VLPs derived from HBcAg or HBsAg, displaying antigens from malaria proteins.

21.5.4 Chemical Conjugation of VLP-Based Assemblies

An alternative approach for displaying antigens on the surface of VLPs is the use of modular systems, in which the native VLP and target antigen are synthesised separately and then chemically conjugated *in vitro*, covalently or noncovalently, linking the antigen to the surface of preassembled VLPs (Fig. 21.7) (see also Chap. 22).

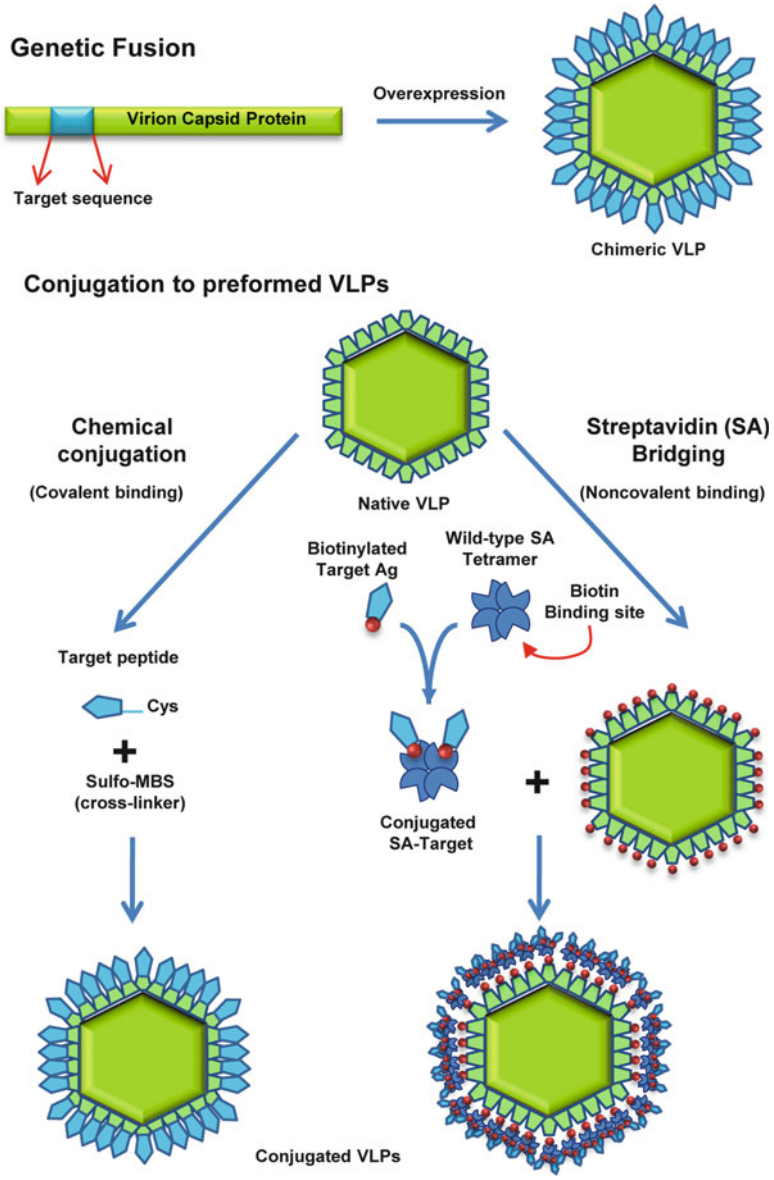


Fig. 21.7 Different approaches for displaying antigens on the surface of VLPs. For the genetic insertion of target sequences, the DNA encoding an antigenic peptide is cloned into the gene encoding the self-assembly-competent polypeptide that forms the VLPs. The resulting construction self-assembles into chimeric VLPs. Chemical conjugation techniques rely on the use of modular systems, in which the native VLP and target antigen are synthesised separately and then conjugated *in vitro*, covalently or noncovalently, linking the antigen to the surface of preassembled VLPs

An advantage of this approach is that the size and structure of the target antigen to be attached are not constrained by the requirements of the correct folding of the VLP monomers and particle assembly. Furthermore, chemical conjugation allows the attachment to the surface of VLPs of diverse kinds of target antigens: short linear peptides, cyclic peptides, full-length proteins, and even nonprotein targets, such as glycans or small haptens (organic molecules that can elicit an immune response only when attached to a protein carrier). A recent example is the antismoking candidate vaccine Nicotine-Qbeta (Cytos Biotechnology), which is produced by covalently coupling nicotine to the surface of VLPs derived from the bacteriophage Qβ. This vaccine is currently in late stages of clinical evaluation and it has been shown to induce high titer nicotine-specific antibody responses in immunised subjects [20, 24].

The conjugation techniques (see also Chap. 22) rely on the presence of addressable moieties on the surface of VLPs, typically amine or sulfhydryl residues. The most commonly used chemical techniques are those employed routinely for protein derivatization [59]: acylation of the amino groups of lysine side chains and the N -terminus, alkylation of the sulfhydryl group of cysteine, and, to a more limited extent, activation of carboxylic acid residues and coupling with added amines. Usually, surface-exposed lysine residues on virus capsid proteins are targeted for conjugation of foreign antigens. Covalent linkage can be achieved by the use of diverse chemical crosslinkers, normally, heterobifunctional conjugation reagents containing two distinct reactive groups that couple to different functional targets, one on the VLP and the other on the antigen. For example, using maleimidobenzoic acid sulfosuccinimidyl ester (Sulfo-MBS), cysteine-containing antigens can be conjugated at high density (*i.e.*, one to three molecules per VLP protein monomer) to exposed amines (lysine residues) on the surface of VLPs [60]. If needed, viruses and VLPs can be engineered to contain useful attachment sites on the surface of the particle [59, 61]. For instance, HBcAg VLPs, bacteriophage MS2 and cowpea mosaic virus have been mutated so that they contain a single surface-exposed reactive residue (lysine or cysteine) per capsid subunit, suitable for antigen conjugation [14, 59, 60]. To ensure that the antigen is coupled to the VLP in a directed and oriented fashion, peptide antigens can be engineered to contain either an amino- or carboxy-terminal amino acid linker sequence containing a free cysteine group.

For noncovalent conjugation of antigens to VLPs, exposed lysine residues can be biotinylated and then attached to biotinylated target antigens through the use of a streptavidin linking molecule [24]. Using this approach VLPs displaying peptides derived from self antigens (*i.e.*, Aβ, CCR-5 or TNF-α) have been obtained [24].

Hence, the flexibility of the chemical conjugation approach offers substantial inherent advantages. However, from the manufacturing point of view, it poses some drawbacks. Chemical linkage results in less than the 100 % of foreign antigen insertion achieved with the genetic fusion method, and is not as reproducible. The production process of conjugated VLPs therefore entails extra challenges and the quality control methods are inevitably more complex.

21.6 Use of Plant and Insect-Derived Chimeric Virions for Foreign Antigen Display

An interesting approach related to the VLP-based vaccine technology is the use of genetically engineered plant viruses as vectors for the presentation of foreign epitopes. In this approach the heterologous peptides are inserted by genetic fusion at exposed loops within viral capsid proteins, but using infectious viral clones for cell transfection, resulting in the formation of viable chimeric virus particles. In contrast with VLPs (in the strict sense frequently given to this definition), these chimeric whole-virus particles are replication-competent and are therefore infectious in plants, their natural hosts. The engineered viruses are propagated in plants and subsequently, the corresponding purified chimeric virions are used as immunogens in target animals. Since the plant viruses are unable to replicate in mammals, these chimeric virions behave similarly to VLPs in terms of safety and the immune response induced. Both, filamentous and icosahedral plant viruses have been successfully developed as foreign epitope presentation systems [62]. The most frequently used systems are cowpea mosaic virus (CPMV), tobacco mosaic virus, cucumber mosaic virus, alfalfa mosaic virus and papaya mosaic virus.

CPMV has been extensively studied for vaccine applications [63–65]. It is a comovirus in the picornavirus superfamily, with a capsid composed of two protein subunits termed small (S) and large (L). The S subunit is about 23 kD and folds into a jellyroll β-sandwich, while the L subunit folds into two jellyroll β-sandwiches, with a total mass of 41 kD. Sixty copies of both proteins form the virus capsid 30 nm particle with a $T = 3$ icosahedral symmetry. Knowledge of the detailed three-dimensional structure of the particle allowed the identification of suitable sites for the insertion of foreign peptides, and the availability of infectious cDNA clones enabled engineering the corresponding genetic changes in the viral genome. Thus, CPMV became the first plant virus to be successfully developed as a peptide presentation platform [65]. In most cases, the foreign sequence has been inserted into the most exposed loop of the virus surface, the βB-βC loop of the S protein, resulting in the display of 60 copies of inserted peptide on the surface of the particle. Other sites, such as the βC′-βC″ loop of the S protein and the βE-αB loop of the L protein, have also been used successfully. Provided the inserted peptide is less than 40 amino acids and has a pI below 9.0, the yields of modified particles are generally similar to those obtained with wild-type CPMV (up to 1 mg of particles per gram of infected leaf tissue).

CPMV particles have been genetically modified to display foreign epitopes derived from mink enteritis virus (MEV) and canine parvovirus (CPV) VP2 proteins, as well as HIV gp41 protein. These chimeric particles were shown to induce neutralising antibody responses when injected into experimental animals, and in the cases of the chimeric particles harbouring epitopes derived from MEV and CPV, the immune responses induced resulted in protection against viral challenges with the corresponding viruses. Likewise chimeric CPMV particles expressing a foreign antigen derived from the outer membrane protein F of *Pseudomonas aeruginosa*

were able to protect mice against challenge by two different immunotypes of *P. aeruginosa* in a model of chronic pulmonary infection.

Biochemical analysis of chimeric CPMV virions with foreign antigens inserted at the βB-βC loop or the βC′-βC″ loop identified a proteolytic cleavage event near the carboxy-terminus of the inserted sequence, which appears to be position, rather than sequence, dependent. This cleavage does not result in the loss of the epitope from the surface of the virion, but as a result the epitope is only anchored to the viral surface by its N-terminus.

An interesting feature of CPMV as an antigen presentation system is the fact that the conditions for crystallization of the virus are well established. Hence, modified particles can readily be crystallised and the structure of the inserted peptide determined by X-ray crystallography, providing a unique opportunity for analysing how the mode of presentation of a peptide affects its immunogenicity. In fact, such a study was conducted by generating a series of chimeric CPMV constructions expressing the 14 amino acid NIm-1A epitope from human rhinovirus 14 (HRV-14), at different positions on the capsid surface [64]. Biochemical and crystallographic analysis of constructions expressing the NIm-1A epitope inserted into the βC′-βC″ loop of the S protein revealed that, although the inserted peptide was free at its C-terminus, it adopted a conformation distinct from that found when a similarly cleaved peptide was expressed in the βB-βC loop of the S protein. Adjustment of the insertion site within the βB-βC loop (by moving it one residue to the left) resulted in the isolation of a chimeric virus in which cleavage at the C-terminus of the foreign epitope was greatly reduced. Crystallographic analysis confirmed that in this case the epitope was presented as a closed loop. Polyclonal antisera raised against this chimeric virus had a significantly enhanced ability to bind to intact HRV-14 particles, compared with antisera raised against other constructs presenting the same epitope as peptides with free C-termini (the antisera raised against these constructions reacted strongly against HRV-14 VP1 in Western blots but bound poorly to intact HRV-14). In conclusion this study demonstrated that the mode of presentation of an epitope on a heterologous carrier can dramatically affect its immunological properties. For epitopes such as NIm-1A, which adopt a constrained structure in their native context (HRV-14 particles), presentation as a closed loop is probably essential for good mimicry. In contrast, the success of the chimeric CPMV particles displaying foreign antigens derived from parvoviruses, or the bacteria *P. aeruginosa* (described above), in stimulating protective immunity, probably stems from the fact that those antigens act as linear epitopes that are active even in a denatured form, because they fold on recognition by the corresponding antibodies and the entropic cost of their folding is not enough to prevent binding with enough affinity. Thus the accurate structural mimicry is not an essential requirement in those cases. Overall, the results of this study demonstrated the potential of the CPMV-based display platform for studying the relationship between peptide structure and immunogenicity.

Whereas no known plant virus can infect mammalian cells, the ability of the chimeric viruses to replicate in plants and spread in the environment, raises biosafety concerns associated with the transport, distribution, and administration

of these plant virus-based vaccines. One option to circumvent these problems is to render the chimeric viruses non-infectious using inactivation methods, such as irradiation with ultraviolet light or chemical treatments. However, these processes have to be carefully monitored as they risk altering the structural properties of the particles and therefore their immunogenic characteristics. Other possibility is to use VLPs derived from the plant virus capsid proteins instead of the infectious chimeric virus particles, for the presentation of the foreign epitopes [66]. Furthermore, since the VLPs do not longer need to be competent at packaging RNA or spreading within plant tissues, as the replication competent chimeric viruses do, this enables to extend the range of mutations (*i.e.*, more insertion sites and longer insertions) that it is possible to introduce into the capsid proteins, thereby extending the versatility of the particles as antigen display systems for vaccine development.

Insect viruses can also be used as platforms for antigen display (they do not replicate in mammalian cells). Flock house virus (FHV) particles are widely used for antigen display, attachment, and delivery in animals [63, 67]. FHV is a member of the insect virus family nodavirus. The capsid (30 nm) is composed of 180 subunits of the single coat protein, arranged with a $T = 3$ icosahedral symmetry. Several surface exposed loops are promising sites for insertion of foreign antigens. FHV particles can be used either as intact virions (and thus, replication competent in insect cells), or as VLPs derived from the expression of the coat protein in the baculovirus expression system.

An interesting application of the FHV-based display platform was the construction of chimeric FHV particles incorporating the extracellular domain of the anthrax toxin receptor 2 (ANTXR2) [67]. The foreign sequence was inserted into the two most exposed loops, at positions 206 and 264, on the FHV coat protein. The extracellular ANTXR2 domain contains 181 amino acids and adopts a compact Rossmann-like α/β fold. Importantly, the termini of this domain are separated by only 4.8 Å in the native structure, representing an ideal situation for insertion into a loop on a carrier protein. Indeed, the chimeric proteins assembled into VLPs. Cryo-EM and three dimensional image reconstruction confirmed that both types of particles displayed new density at higher radius. Pseudoatomic models of the particles generated by docking the X-ray coordinates of the FHV coat protein and ANTXR2 domain into the cryo-EM density revealed the expected differences in the geometric pattern in which the ANTXR2 domains are displayed on the surface. For insertion site 206, the ANTXR2 domains were clustered in groups of six at the twofold axes of the particle, whereas insertion at site 264 resulted in more even distribution of the foreign domain. The accurate folding of the inserted protein was confirmed by demonstrating that the chimeric particles functioned as an anthrax antitoxin *in vitro* and *in vivo*. Specifically, like native ANTXR2, the particles were capable of binding anthrax protective antigen (PA), which forms part of anthrax lethal toxin and edema toxin. Based on this ability, the chimeric particles could potentially be used as a therapeutic compound to treat anthrax infections. Interestingly, because of the differences in ANTXR2 display pattern, the two types of VLPs showed different potencies as an antitoxin, with chimera 264 having a lower IC_{50} for toxin neutralization than chimera 206. Computational modelling suggested that

this was because chimera 264 bound more PA molecules than chimera 206. Although both particles displayed 180 ANTXR2 domains, steric hindrance prevented full occupancy of these ligands. Instead, it was predicted that chimera 264 might bind 120–130 PA molecules whereas chimera 206 could only bind 60–90 PA molecules.

Given that the binding of PA to ANTXR2 is exceptionally strong (dissociation constant $K_D = 170$ pM) complexes formed between the chimeric particles and PA can be expected to be very stable. This prompted immunogenicity studies based on the assumption that polyvalent display of PA would induce a more potent immune response to this antigen than monomeric, recombinant PA, which is being developed as a second generation anthrax vaccine. Indeed, rats survived lethal toxin challenge 4 weeks after a single immunization with the VLP 264-PA complex, whereas animals injected with an equivalent amount of recombinant PA died. This result reflects rapid production of neutralising antibodies in the absence of an adjuvant, two key goals for the development of an improved anthrax vaccine. The chimeric FHV particles thus, might serve a dual purpose in functioning as an anthrax toxin inhibitor and in forming a basis for development of a new anthrax vaccine.

21.7 Perspectives and Conclusions

VLPs are appealing as vaccine candidates because their inherent properties (*i.e.*, virus-sized, multimeric antigens, highly organised and repetitive structure, not infectious) make them suitable for the induction of safe and efficient humoral and cellular immune responses. Currently, there is a clear trend towards the establishment of VLPs as a powerful tool for vaccine development. VLP-based vaccines have already been licensed for human diseases (HBV and HPV) as well as for use in the veterinary field (PCV2), and many more vaccine candidates are currently in late stages of evaluation. Moreover, the development of VLPs as platforms for foreign antigen display has further broadened their potential applicability both as prophylactic and therapeutic vaccines.

Currently, main efforts in VLP technology are focused in the development of new VLP-based antigen display platforms for vaccine development. Structural characterisation of VLPs is an important requirement for this aim. Structural studies are mainly performed by advanced cryo-EM and X-ray crystallography, focusing on the comparison of VLPs composed of different numbers and combinations of structural proteins. The use of biochemical methods provides details on individual viral structural components, as well as insights into the structural basis of assembly, packaging and the interactions of VLPs with host components. Many studies are aimed at characterising the minimal requirements for VLP formation or contemplate prospects of modifying the original proteins without hampering the natural ability of these proteins to assemble into highly organised macromolecules. As a

consequence, modified structural proteins appropriate for assembly of multipurpose chimeric VLPs can be eventually designed.

The relative ability of diverse VLP types to induce the different branches of the immune response is influenced by a number of factors that are VLP-specific. Therefore, it appears unlikely that a single VLP platform will meet all the desired requirements. However, the continued parallel development of multiple VLP platforms will ensure that individual vaccines can be tailored appropriately to the type of immune response required in each case.

Acknowledgements We are indebted to current and former members of our group (Iván Angulo, Carolina Cubillos, Horacio Almanza, Noelia Moreno, Ignacio Mena, Beatriz Guerra and Yolanda Gómez), and other collaborators: Elisa Crisci and María Montoya (CReSA), José R. Castón (CNB-CSIC), for their work and skills, that made our work in this field possible. Our work is funded by grants from the Spanish Ministry of Science and Innovation: AGL2010-22200-C02, CSD 2006-00007 (PORCIVIR, program CONSOLIDER-INGENIO 2010), and EU: NADIR-UE-228394.

References and Further Reading

1. Meeusen EN, Walker J, Peters A, Pastoret PP, Jungersen G (2007) Current status of veterinary vaccines. Clin Microbiol Rev 20:489–510
2. Plotkin SA, Plotkin SL (2011) The development of vaccines: how the past led to the future. Nat Rev Microbiol 9:889–893
3. Levine MM, Dougan G, Good MF, Liu MA, Nabel GJ, Nataro JP, Rapuoli R (2009) New generation vaccines, 4th edn. Informa Healthcare, New York
4. Kaufmann SHE (2004) Novel vaccination strategies. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
5. Plotkin S (2011) History of vaccine development. Springer, New York
6. Murdin AD, Barreto L, Plotkin S (1996) Inactivated poliovirus vaccine: past and present experience. Vaccine 14:735–746
7. Cottam EM, Wadsworth J, Shaw AE, Rowlands RJ, Goatley L, Maan S, Maan NS, Mertens PP, Ebert K, Li Y, Ryan ED, Juleff N, Ferris NP, Wilesmith JW, Haydon DT, King DP, Paton DJ, Knowles NJ (2008) Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. PLoS Pathog 4:e1000050
8. Buonaguro L, Tornesello ML, Buonaguro FM (2010) Virus-like particles as particulate vaccines. Curr HIV Res 8:299–309
9. Roldao A, Mellado MC, Castilho LR, Carrondo MJ, Alves PM (2010) Virus-like particles in vaccine development. Expert Rev Vaccines 9:1149–1176
10. Grgacic EV, Anderson DA (2006) Virus-like particles: passport to immune recognition. Methods 40:60–65
11. Jansen KU, Conner ME, Estes MK (2009) Virus-like particles as vaccines and vaccine delivery systems. In: Levine MM, Dougan G, Good MF, Liu MA, Nabel GJ, Nataro JP, Rapuoli R (eds) New generation vaccines. Informa Healthcare, New York, USA, pp 298–305
12. Crisci E, Barcena J, Montoya M (2012) Virus-like particles: the new frontier of vaccines for animal viral infections. Vet Immunol Immunopathol 148:211–225
13. Roy P, Noad R (2009) Virus-like particles as a vaccine delivery system: myths and facts. Adv Exp Med Biol 655:145–158
14. Chackerian B (2007) Virus-like particles: flexible platforms for vaccine development. Expert Rev Vaccines 6:381–390

15. Bachmann MF, Jennings GT (2004) Virus-Like particles: combining innate and adaptive immunity for effective vaccination. In: Kaufmann SHE (ed) Novel vaccination strategies. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 415–432
16. Buonaguro L, Tagliamonte M, Tornesello ML, Buonaguro FM (2011) Developments in virus-like particle-based vaccines for infectious diseases and cancer. Expert Rev Vaccines 10:1569–1583
17. Plummer EM, Manchester M (2011) Viral nanoparticles and virus-like particles: platforms for contemporary vaccine design. Wiley Interdiscip Rev Nanomed Nanobiotechnol 3:174–196
18. Dyer MR, Renner WA, Bachmann MF (2006) A second vaccine revolution for the new epidemics of the 21st century. Drug Discov Today 11:1028–1033
19. Jennings GT, Bachmann MF (2009) Immunodrugs: therapeutic VLP-based vaccines for chronic diseases. Annu Rev Pharmacol Toxicol 49:303–326
20. Bachmann MF, Jennings GT (2011) Therapeutic vaccines for chronic diseases: successes and technical challenges. Philos Trans R Soc Lond B Biol Sci 366:2815–2822
21. Zepp F (2010) Principles of vaccine design-lessons from nature. Vaccine 28(Suppl 3): C14–C24
22. Jennings GT, Bachmann MF (2007) Designing recombinant vaccines with viral properties: a rational approach to more effective vaccines. Curr Mol Med 7:143–155
23. Spohn G, Bachmann MF (2008) Exploiting viral properties for the rational design of modern vaccines. Expert Rev Vaccines 7:43–54
24. Jennings GT, Bachmann MF (2008) The coming of age of virus-like particle vaccines. Biol Chem 389:521–536
25. Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2003) Virus offence meets host defense. In: Principles of virology: molecular biology, pathogenesis, and control of animal viruses, 2nd edn. ASM Press, Washington, DC, pp 531–584
26. Bachmann MF, Jennings GT (2010) Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. Nat Rev Immunol 10:787–796
27. Xiang SD, Scholzen A, Minigo G, David C, Apostolopoulos V, Mottram PL, Plebanski M (2006) Pathogen recognition and development of particulate vaccines: does size matter? Methods 40:1–9
28. Scheerlinck JP, Greenwood DL (2008) Virus-sized vaccine delivery systems. Drug Discov Today 13:882–887
29. De Temmerman ML, Rejman J, Demeester J, Irvine DJ, Gander B, De Smedt SC (2011) Particulate vaccines: on the quest for optimal delivery and immune response. Drug Discov Today 16:569–582
30. Bachmann MF, Rohrer UH, Kundig TM, Burki K, Hengartner H, Zinkernagel RM (1993) The influence of antigen organization on B cell responsiveness. Science 262:1448–1451
31. Hinton HJ, Jegerlehner A, Bachmann MF (2008) Pattern recognition by B cells: the role of antigen repetitiveness *versus* Toll-like receptors. Curr Top Microbiol Immunol 319:1–15
32. Bachmann MF, Zinkernagel RM (1997) Neutralizing antiviral B cell responses. Annu Rev Immunol 15:235–270
33. Larsson M, Beignon AS, Bhardwaj N (2004) DC-virus interplay: a double edged sword. Semin Immunol 16:147–161
34. Moron G, Dadaglio G, Leclerc C (2004) New tools for antigen delivery to the MHC class I pathway. Trends Immunol 25:92–97
35. Lenz P, Day PM, Pang YY, Frye SA, Jensen PN, Lowy DR, Schiller JT (2001) Papillomavirus-like particles induce acute activation of dendritic cells. J Immunol 166:5346–5355
36. Gedvilaite A, Dorn DC, Sasnauskas K, Pecher G, Bulavaite A, Lawatscheck R, Staniulis J, Dalianis T, Ramqvist T, Schonrich G, Raftery MJ, Ulrich R (2006) Virus-like particles derived from major capsid protein VP1 of different polyomaviruses differ in their ability to induce maturation in human dendritic cells. Virology 354:252–260
37. Johnson JE, Chiu W (2000) Structures of virus and virus-like particles. Curr Opin Struct Biol 10:229–235

38. Brun A, Barcena J, Blanco E, Borrego B, Dory D, Escribano JM, Le Gall-Recule G, Ortego J, Dixon LK (2011) Current strategies for subunit and genetic viral veterinary vaccine development. Virus Res 157:1–12
39. Schneider-Ohrum K, Ross TM (2012) Virus-like particles for antigen delivery at mucosal surfaces. Curr Top Microbiol Immunol 354:53–73
40. Garcea RL, Gissmann L (2004) Virus-like particles as vaccines and vessels for the delivery of small molecules. Curr Opin Biotechnol 15:513–517
41. Roldao A, Silva AC, Mellado MCM, Alves PM, Carrondo MJT (2011) Viruses and virus-like particles in biotechnology: fundamentals and applications. In: Moo Y (ed) Comprehensive biotechnology, vol. 1: Scientific fundamentals in biotechnology, 2nd edn. Elsevier/Pergamon, Oxford, pp 625–649
42. Pumpens P, Ulrich RG, Sasnauskas K, Kazaks A, Ose V, Grens E (2008) Construction of novel vaccines on the basis of virus-like particles: hepatitis B virus proteins as vaccine carriers. In: Khudyakov YE (ed) Medicinal protein engineering. CRC Press, Boca Raton, Florida, USA, pp 205–247
43. Kirnbauer R, Booy F, Cheng N, Lowy DR, Schiller JT (1992) Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc Natl Acad Sci U S A 89:12180–12184
44. Campo MS, Roden RB (2010) Papillomavirus prophylactic vaccines: established successes, new approaches. J Virol 84:1214–1220
45. Moser C, Amacker M, Zurbriggen R (2011) Influenza virosomes as a vaccine adjuvant and carrier system. Expert Rev Vaccines 10:437–446
46. Cox MM (2012) Recombinant protein vaccines produced in insect cells. Vaccine 30:1759–1766
47. Vicente T, Roldao A, Peixoto C, Carrondo MJ, Alves PM (2011) Large-scale production and purification of VLP-based vaccines. J Invertebr Pathol 107(Suppl):S42–48
48. Casal JI (2001) Use of the baculovirus expression system for the generation of virus-like particles. Biotechnol Genet Eng Rev 18:73–87
49. Vicente T, Mota JP, Peixoto C, Alves PM, Carrondo MJ (2011) Rational design and optimization of downstream processes of virus particles for biopharmaceutical applications: current advances. Biotechnol Adv 29:869–878
50. Pattenden LK, Middelberg AP, Niebert M, Lipin DI (2005) Towards the preparative and large-scale precision manufacture of virus-like particles. Trends Biotechnol 23:523–529
51. Liu W, Chen YH (2005) High epitope density in a single protein molecule significantly enhances antigenicity as well as immunogenicity: a novel strategy for modern vaccine development and a preliminary investigation about B cell discrimination of monomeric proteins. Eur J Immunol 35:505–514
52. Pumpens P, Grens E (2001) HBV core particles as a carrier for B cell/T cell epitopes. Intervirology 44:98–114
53. Lee KK, Johnson JE (2003) Complementary approaches to structure determination of icosahedral viruses. Curr Opin Struct Biol 13:558–569
54. Arnon R, Van Regenmortel MH (1992) Structural basis of antigenic specificity and design of new vaccines. FASEB J 6:3265–3274
55. Casal JI, Rueda P, Hurtado A (1999) Parvovirus-like particles as vaccine vectors. Methods 19:174–186
56. Rueda P, Hurtado A, del Barrio M, Martinez-Torrecuadrada JL, Kamstrup S, Leclerc C, Casal JI (1999) Minor displacements in the insertion site provoke major differences in the induction of antibody responses by chimeric parvovirus-like particles. Virol 263:89–99
57. Billaud JN, Peterson D, Barr M, Chen A, Sallberg M, Garduno F, Goldstein P, McDowell W, Hughes J, Jones J, Milich D (2005) Combinatorial approach to hepadnavirus-like particle vaccine design. J Virol 79:13656–13666
58. Greenstone HL, Nieland JD, de Visser KE, De Bruijn ML, Kirnbauer R, Roden RB, Lowy DR, Kast WM, Schiller JT (1998) Chimeric papillomavirus virus-like particles elicit antitumor

immunity against the E7 oncoprotein in an HPV16 tumor model. Proc Natl Acad Sci U S A 95:1800–1805
59. Strable E, Finn MG (2009) Chemical modification of viruses and virus-like particles. Curr Top Microbiol Immunol 327:1–21
60. Jegerlehner A, Tissot A, Lechner F, Sebbel P, Erdmann I, Kundig T, Bachi T, Storni T, Jennings G, Pumpens P, Renner WA, Bachmann MF (2002) A molecular assembly system that renders antigens of choice highly repetitive for induction of protective B cell responses. Vaccine 20:3104–3112
61. Mateu MG (2011) Virus engineering: functionalization and stabilization. Protein Eng Des Sel 24:53–63
62. Yusibov V, Rabindran S, Commandeur U, Twyman RM, Fischer R (2006) The potential of plant virus vectors for vaccine production. Drugs R D 7:203–217
63. Destito G, Schneemann A, Manchester M (2009) Biomedical nanotechnology using virus-based nanoparticles. Curr Top Microbiol Immunol 327:95–122
64. Steinmetz NF, Lin T, Lomonossoff GP, Johnson JE (2009) Structure-based engineering of an icosahedral virus for nanomedicine and nanotechnology. Curr Top Microbiol Immunol 327:23–58
65. Sainsbury F, Canizares MC, Lomonossoff GP (2010) Cowpea mosaic virus: the plant virus-based biotechnology workhorse. Annu Rev Phytopathol 48:437–455
66. Saunders K, Sainsbury F, Lomonossoff GP (2009) Efficient generation of cowpea mosaic virus empty virus-like particles by the proteolytic processing of precursors in insect cells and plants. Virol 393:329–337
67. Venter PA, Schneemann A (2008) Recent insights into the biology and biomedical applications of Flock House virus. Cell Mol Life Sci 65:2675–2687

Further Reading

Especially recommended for further reading are references [9, 13, 14, 17, 24, 26, 41, 42, 64, 65] *listed above.*

Chapter 22
Nanoscale Science and Technology with Plant Viruses and Bacteriophages

Alexander M. Bittner, José María Alonso, Marcin Ł. Górzny, and Christina Wege

Abstract Nanoscale science refers to the study and manipulation of matter at the atomic and molecular scales, including nanometer-sized single objects, while nanotechnology is used for the synthesis, characterization, and for technical applications of structures up to 100 nm size (and more). The broad nature of the fields encompasses disciplines such as solid-state physics, microfabrication, molecular biology, surface science, organic chemistry and also virology. Indeed, viruses and viral particles constitute nanometer-sized ordered architectures, with some of them even able to self-assemble outside cells. They possess remarkable physical, chemical and biological properties, their structure can be tailored by genetic engineering and by chemical means, and their production is commercially viable. As a consequence, viruses are becoming the basis of a new approach to the manufacture of nanoscale materials, made possible only by the development of imaging and manipulation techniques. Such techniques reach the scale of single molecules and nanoparticles. The most important ones are electron microscopy and scanning probe microscopy (both awarded with the Nobel Prize in Physics 1986 for the engineers and scientists who developed the respective instruments). With nanotechnology being based more on experimental than on theoretical investigations, it emerges that physical virology can be seen as an intrinsic part of it.

A.M. Bittner (✉)
CIC nanoGUNE, Av. Tolosa 76, 20018 San Sebastián, Spain

Ikerbasque, Basque Foundation for Science, 48011 Bilbao, Spain
e-mail: a.bittner@nanogune.eu

J.M. Alonso • M.Ł. Górzny
CIC nanoGUNE, Av. Tolosa 76, 20018 San Sebastián, Spain

C. Wege
Institute of Biology, University of Stuttgart, Pfaffenwaldring 57, D-70550 Stuttgart, Germany
e-mail: christina.wege@bio.uni-stuttgart.de

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8_22,

Keywords Bacteriophage • Biophysics • Capsid • DNA • Drug delivery • Hybrid materials • Lithography materials • Microscopy • RNA • Nanoscience • Nanotechnology • Scanning probe • Template • Virus

Abbreviations

1D	One-dimensional
3D	Three-dimensional
AFM	Atomic force microscopy
CCMV	Cowpea chlorotic mottle virus
CP	Capsid protein
CPMV	Cowpea mosaic virus
DEP	Dielectrophoresis
eBL	Electron beam lithography
GFP	Green fluorescent protein
ORF	Open reading frame
SEM	Scanning electron microscopy
SPM	Scanning probe microscopy
STM	Scanning tunneling microscopy
TEM	Transmission electron microscopy
TMV	Tobacco mosaic virus
VLP	Virus-like particles

22.1 Introduction: Viral vs. Artificial (Synthetic) Nanostructures

Synthetic nanoscale structures are closely linked to the subject of *nanotechnology*. This field, often also identified with *nanoscale science*, deals with the production and manipulation of matter at the atomic, molecular, and supramolecular levels. It covers the length scales from 1 to 100 nm (in an extended definition 0.1–1000 nm). Nanotechnology provides a fundamental understanding of phenomena and materials at the nanoscale, with the aims of using structures, devices and systems that have novel properties and functions, due to their small size. It involves also controlling the structures, and integrating them into micro- and macroscale material components, systems and architectures. Within these larger and more complex assemblies, the control and construction of the underlying structures and components remains at the nanometer scale. A similar organization is found in cells, which operate on the microscale, but rely entirely on the interplay of nanostructures such as proteins. In extreme cases, the critical length scale for novel properties and phenomena may be <1 nm (*e.g.*, for the manipulation of single atoms in scanning probe techniques),

or >100 nm (*e.g.*, nanoparticle-reinforced polymers can exhibit features of >200 nm size, based on the local bonds between the particles and the polymer).

Nanotechnology is a very broad concept, which implies the application of fields of science as diverse as surface science, organic chemistry, molecular biology, semiconductor and solid-state physics, microfabrication, etc. A huge driving force for a proper technology is that materials can effectively be made to be stronger, lighter, more durable, more reactive, more porous, or more conductive, among many other properties. However, the concept encompasses not only very recent, but also some very well established processes and products. Hence, hundreds of everyday commercial products rely on nanoscale materials and processes, from paint to computer chips. Many more will follow, so nanotechnology is a key technology for the future, and governments have invested billions of dollars in research.

Focusing on nanotechnology on the molecular scale of 1–10 nm, very few products are developed, the probably best-known being based on gold particles. A crucial issue on this scale is the development of simple construction schemes for the mass fabrication of identical nanoscale structures, just like chemical reactions can build molecules. Conventional "top-down" fabrication techniques rely on demagnification of a given structure; this approach can be energy-intensive and wasteful: Many production steps involve depositing unstructured layers and then patterning them by removing most of the deposited material. Moreover, increasingly expensive fabrication facilities are required as the feature size decreases. The natural alternative to "top-down" construction is the "bottom-up" approach. In this case nanoscale structures are built from their atomic and molecular constituents by self-assembly. However, the "bottom-up" methodology is technologically not yet possible because the assembly processes are slow, faulty and in most cases not sufficiently well controlled. Organic chemists try to tackle this problem by developing ever more complex subunits that assemble into nanostructures; on the other hand, specific intermolecular interactions and tailored self-assembly come for free with natural building blocks, such as virus capsid proteins, which are able to spontaneously assemble, under the proper conditions, into virus capsids and virus-like particles (VLPs) (see Chaps. 2, 10, 11, 19 and 21). Therefore, virus structures can and are being used as templates for directing the self-assembly of materials, and as scaffolds for nanofabrication processes.

A completely different motivation for the biomolecular approach [1] is based on the prefix "bio": The very fact that biomolecules are molecules of life implies that they have numerous active functional moieties that interact in complex ways with their environment. This can be used for binding or for the synthesis of inorganic and organic substances. Chemically reactive moieties on the biomolecules, such as amines and carboxyl groups, can be exploited to attract and react with other molecules. Second, substrate-specific affinities (antibody-antigen, biotin-streptavidin, oligonucleotide base pairs) have been employed to assemble substances in a programmed position, to align structured materials in a specially designed pattern, and to conjugate biomolecular substances with each other. Third, the

enzymatic activity of biomolecules is utilized to decompose or generate organic and inorganic substances by stabilizing intermediates during reactions.

A variety of biomolecules have already been exploited for the preparation of nanoscale materials. Oligonucleotides are commonly used in bionanotechnology due to their hybridizing functions and their ability to create reconfigurable structures [2]. Peptides can be synthesized in the laboratory and conjugated with organic molecules while maintaining the biological activities of catalysis and specific recognition. Certain peptides are able to mineralize inorganic sources with functionalities that are dependent on the composition and the structure of the peptide. Proteins and their assemblies serve as platform for nanomaterials synthesis as well. Proteins display secondary molecular forces such as hydrogen bonding, electrostatic and hydrophobic interactions that play an essential role in scaffolds for the fabrication of nanomaterials with defined geometry. Viruses are another type of biological macromolecular complexes that may be used for those purposes. The application of animal-infecting virus species and derivatives thereof is centered mainly on biomedical problems such as the development of antiviral agents and vaccines (see Chaps. 20 and 21) (with some notable exceptions); reasons for the infrequent use of animal viruses in nanotechnology include the possible pathogenicity of complete viruses, their dependance on special cells for their propagation, lower yields, greater expense and, frequently, high structural complexity with outer lipid envelopes. In contrast, plant and bacterial viruses (bacteriophages) are suitable for a much wider range of nanotechnological purposes because they have several advantageous characteristics [1, 3]. Major features that make virus particles potentially useful for many nanotechnological applications include:

1. Viral particles possess precise nanoscale structure and dimensions, in contrast to artificial nanomaterials prepared by "top-down" approaches. Their size range, from about 15 nm up to >2 μm (in diameter or length), is unique for organic structures, and in some cases characterized down to atomic resolution. Synthetic colloids and polymers of comparable dimensions rarely show such geometrical and chemical precision.
2. Viruses can be found in a variety of shapes (most importantly, icosahedron-derived, *i.e.* 'spherical' symmetries, and helical shapes (tubes and filaments), but also bullet-shaped particles; see Chap. 2). For most types of plant viruses and bacteriophages, all viral particles are identical in size and composition (in some cases the capsid exists in more than one size or shape), with one or more nucleic acid molecules (DNA or RNA) packaged into a protein shell made up of numerous subunits of one or a low number of distinct protein types. (A few plant viruses also produce empty additional capsids, and some harbour an outer lipid layer).
3. Viral capsids exhibit constrained internal cavities that are accessible to small molecules, but often impermeable to large ones, providing opportunities for assembly and packaging of cargos. The exterior and interior interfaces of the capsid have been utilized for directing encapsulation and synthesis of both inorganic and organic materials. E.g., nanoparticles of transition metal

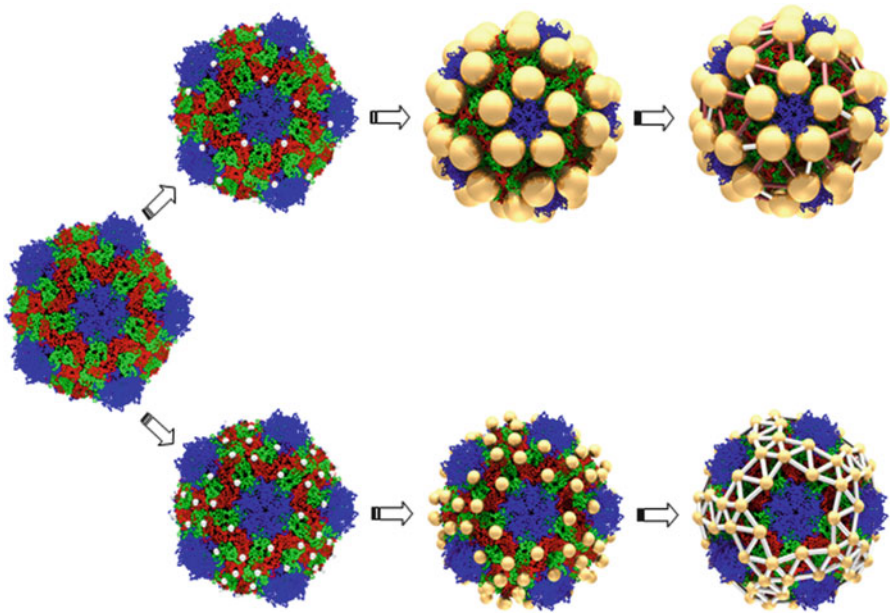


Fig. 22.1 Nanoparticles nucleating on nucleic acid-free CPMV. The *top row* shows a Cys (*white dots*) mutant of CPMV with attached 5 nm gold nanoparticles, and finally interconnected gold particles, bridged by organic molecules. The *bottom row* shows a double Cys mutant (two thiol groups per capsid protein), to which 2 nm gold particles bind, again to be interconnected (reproduced from [5] with permission)

polyoxometalates were nucleated on the interior interface of nucleic acid-free cowpea chlorotic mottle virus (CCMV), and gold particles on cowpea mosaic virus (CPMV) [4] (see Fig. 22.1). Other examples are medically relevant molecules, such as doxorubicin, which have been encapsulated through covalent attachment on the interior surface of virus-like particles from rotavirus structural protein VP6 [6].

4. Viruses represent the best examples of polyvalence and self-assembly processes. Several plant viruses do not depend on living cells for their self-assembly, but can be reconstituted from isolated components also *in vitro*. Viral particles have large surface areas, which permits the display of many molecules without concern of steric hindrance. Functional groups are displayed in multiple copies on the coat (capsid) Protein (CP) subunits forming the viral capsids, and offer anchor points for biochemical interactions.
5. The high strength of capsid protein-protein interactions makes several types of virus particles very stable. They are usually more resistant toward changes of pH, temperature, ionic strength and solvent than standard proteins, and afford a broader range of conditions for their isolation, storage and use.

6. Virus composition and properties may be tuned by manipulation of the viral genome. Exogenous oligopeptide sequences can be inserted with standardized mutagenesis protocols [7] (see also Chap. 21).
7. Mass production of viruses or their components may be easily achieved by growing them in infected hosts or in heterologous *in vitro* production cultures [8, 9] (plants, cultured insect, yeast or bacterial cells) with a yield in the range of 0.1–1 % by weight for certain viruses in suitable production systems. Moreover, viruses exhibit unique densities, making purification steps simpler and faster than those required for most proteins, and thus adjustable to large scale.

From a rather biochemical focus on viral surfaces and their functionalization, the discussion will move to assembly of viruses and of materials (for which viruses are used as templates). The characterization techniques and devices are more based on physics, while biosensors combine all disciplines.

22.2 The Control of Surface Chemistry by Genetic Engineering and by Chemical Reactions

As outlined above and different from inorganic templates, the protein surfaces of robust virus and bacteriophage nanoparticles harbour an enormous potential for the site-directed introduction of novel functional groups and, concomitantly or subsequently, even complex activities [10]. This may (1) modulate the structures' electrostatic charge, adapting them to different technical environments; (2) add novel chemically addressable anchors for the deposition of inorganic compounds such as metal or metal oxides or the immobilization of larger molecules for catalytic or detection purposes; and (3) incorporate extended amino acid sequences functional by themselves, *e.g.* as capture moieties or enzymatic units. The different strategies will be explained and exemplified in the following, with the term 'virus' being used for both plant and bacterial infectious agents unless otherwise stated. Some arguments also hold true for certain animal viruses such as adeno- and adeno-associated viruses (see Chaps. 10 and 11).

22.2.1 Chemical Modifications Using Naturally Occurring Reactive Groups

Among the natural amino acid building blocks of proteins, those with polar, basic or acidic or chemically highly reactive side chains are of superior importance for technical applications of viral templates, if accessible on their outer or interior surfaces – depending on the intended use. Cavities inside viral capsids are frequently exploited as 'casting moulds' to shape and protect inorganic deposits. Nanowires or granules can be formed inside tubular virions such as tobacco mosaic

virus (TMV), or in the shells of spherical (or quasi-icosahedral) viruses, *e.g.* CCMV, red clover necrotic mosaic virus, CPMV or brome mosaic virus, respectively. Additionally, hollow viral protein structures, with or without nucleic acid content, are explored for their potential to serve as vehicles ('capsules' or 'cages') for the targeted delivery of different substances to specific sites both in technical and biological environments (see Sects. 22.6 and 22.8). These applications frequently depend on the charge of the interior protein surfaces, providing attractive electrostatic forces and redox conditions, which retain or precipitate materials of interest inside the nanocontainers. Some strategies also involve direct covalent linkage of target molecules to chemically addressable groups in the viral container, predominantly amino, thiol, and carboxyl functions (see below). Similar prerequisites exist if viruses are applied as 'positive templates', *i.e.* proteinaceous carrier complexes for the fabrication of coated hybrid nanorods, filaments or beads. Depending on the molecules to be exposed on the viral backbones, either noncovalent interactions or chemical binding may be selected to interconnect the viral and the functional component. These two essentially different strategies are suited to mediate hierarchical assembly of virus-containing complex architectures and materials, and to integrate viruses and their derivatives into technical devices (see Sects. 22.4, 22.7 and 22.8).

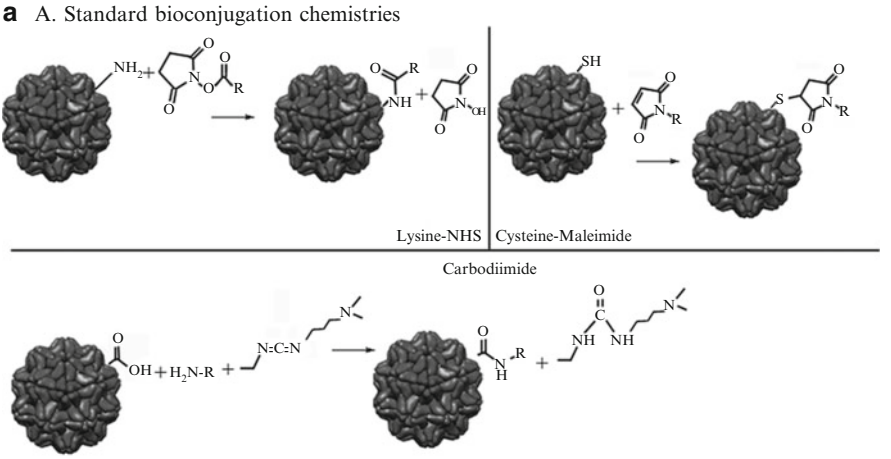
Several naturally existing viral scaffolds of different shapes have been employed successfully for the attachment of inorganic, synthetic or biological molecules to their outer capsid or inner cavities. The reductive (electroless) deposition of metals and metal oxides from ionic precursor salt solutions, for example, depends primarily on the local chemistry of the capsid protein surface. It came out to work well with different, readily available viral templates. Similarly, mineralization of distinct viral capsids with silica coatings succeeded without a need for specifically addressable nucleation sites. Selective and direct covalent fixation of compounds, however, relies on the presence of readily accessible anchor groups. Various viral templates offer amino functions of lysine (Lys) residues, which can be conjugated efficiently with target molecules *via* alkylation or acylation, *e.g.* using N-hydroxysuccinimidyl ester or isothiocyanate derivatives. Thiol groups of cysteine (Cys) are well-established reaction partners for the coupling of sulfhydryl-carrying molecules or alkylation with maleimides [11, 12], but not frequently exposed on viral capsids. Carboxyl functions of acidic amino acids (aspartic and glutamic acid, Asp and Glu) are less attractive reaction partners, since the respective coupling procedures can suffer from the presence of an aqueous environment, which is typically preferred with viral particles, though not essential in any case. Carboxyl derivatization typically makes use of an activation step involving EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) and subsequent reaction of the resulting intermediate with primary amines to form an amide bond. A few other amino acids found on viral capsid surfaces as well as post-translational carbohydrate modifications linked to protein or lipid components of special viruses are, in principle, directly addressable by conjugation techniques as well, but are by far less relevant for straightforward chemical functionalization of virus templates. Finally, aromatic tyrosine (Tyr) side chains can be exploited for sophisticated coupling chemistry "beyond the labeling kit", such as

covalent diazonium linkage, but are more frequently used for intermediate chemical conversions preceding bio-orthogonal coupling (see below, and Fig. 22.2a for most common conjugation reactions to viral proteins). In summary, some natural anchor groups are found and easily used on a number of technically and scientifically interesting viral scaffolds, but many routes of research and development demand additional selectively addressable attachment sites for functional molecules, which need to be generated by biochemical or genetic engineering, respectively, as explained in the following.

22.2.2 Biochemical and Biotechnological Engineering Conferring Novel Addressability

The first and relatively fast biochemical approach to viral capsid functionalization transforms surface-exposed side chains or groups accessible at the ends of CP subunits into advantageous functions. Here, besides Lys and Cys, Tyr with its phenolic hydroxyl group is among the most attractive target amino acids, since it is amenable to both oxime condensation and 'Click' chemistry. The respective two- or multistep reactions equip viral scaffolds with anchor groups or 'handles' accessible to selective, covalent and bioorthogonal coupling mechanisms, which do not affect the integrity of the targeted biomolecule. The most prominent examples are the copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition (CuAAC) 'click' reactions, by use of which a large variety of distinct compounds ranging from fluorescent dyes to complete proteins has been interlinked with viral templates. The second experimental route towards a targeted integration of novel functionalities into virus-like particles is gene technology-based protein modification, that adds numerous further degrees of freedom to the development of virus-based tools. Although it is by far more time-consuming and might bear a substantial risk of failure, genetic engineering has proven the most flexible instrument for equipping viral shells with a multitude of distinct components, by itself or in combination with chemical coupling. Thereby attached functions include recognition and reporter elements as well as catalytically active units (see below). The respective approaches comprise both replacement and insertion of CP amino acid residues to modify local charges, to introduce chemically addressable sites (as described above), and to add peptides or extended protein domains. The following paragraphs give a short overview on essential molecular biology strategies underlying the genetic tailoring of viral capsids or virus-derived nanostructures.

Prerequisite is the availability of the full genetic sequence information of a robust and suitably shaped template virus, or at least of its structural components if they can be assembled to VLPs outside the natural host. Typically, purified viral nucleic acids genomes or cDNA copies thereof have been cloned into bacterial plasmids (see Fig. 22.2b) or other vectors by biotechnology methods involving enzymatic modification, separation, ligation, and amplification steps. After



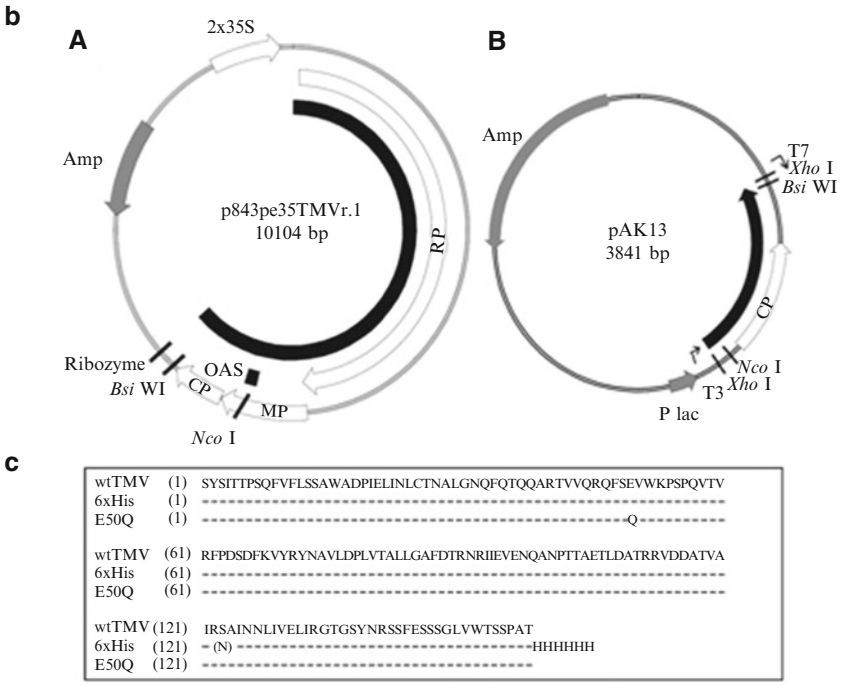


Fig. 22.2 Modification of virus capsids. (**a**) Common bioconjugation reactions: (A) addressing amino, thiol, or carboxyl groups (EDC – 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide); (B) bio-orthogonal coupling by 'click' chemistry on azide-modified amino acids (*e.g.* Lys, Cys or Tyr), reaction of cargo "R" bearing an alkyne (Adapted from [13]). (**b**) Genetic maps of: (A) a plant-infectious TMV plasmid; (B) a small intermediate construct used for mutagenesis of the TMV CP open reading frame; (**c**) alignment of three types of TMV CP amino acid sequences produced in plants: wtTMV (wildtype), 6xHis (C-terminally extended by hexahistidine peptide), E50Q (affecting TMV tube stability). For details, refer to [14]

sequencing has verified the identity of the genetic material, either whole viral genome clones are constructed and multiplied *via* plasmids adapted to infect suitable production hosts (*i.e.* plants or bacteria), or partial sequences encoding the viral CPs are inserted into protein production plasmids (in cases where capsids can be formed in the absence of infectious viral nucleic acid). Most viruses accessible that way form stable virions of relatively low complexity, some of which can also exist as empty shells or assemble *in vitro*. Their genomes are rather small, and for infectious clones, transmission methods for the propagation in plants have been developed, which circumvent the problem that most viruses naturally depend on certain insect, fungal or other vectors for being spread into new plants. Capsids which can be reconstituted *in vitro* may also be assembled from proteins produced in heterologous cell cultures (*e.g.,* bacteria or yeast). While the self-assembly of spherical or quasi-icosahedral viruses mostly does not allow extensive changes of the outer capsid dimensions, tubular or filamentous plant viruses can be shortened *in vitro*, and elongated both *in vitro* and *in vivo* by means of altered encapsidated RNA, which determines the length of the final nucleoprotein particle. Filamentous bacteriophages, though, usually cannot assemble outside cells and thus offer less possibilites of modifying their overall shape. The major aim of genetic engineering approaches, however, is to alter the surface chemistry of the resulting nanoparticles. In this regard, filamentous phages such as M13 or fd harbor superior degrees of freedom in comparison to most plant viruses: They are composed of typically five different proteins encoded independently on a small genome, all of which can be genetically tailored in parallel to change domains at both ends of the viral particles and their longitudinal outer surface selectively. For plant viruses, comparable numbers of different genetically programmed modifications, carried out at pre-determined sites on a single particle, have not been achieved so far. Notwithstanding, *in vitro* particle reconstitution from different types of protein monomers are distinct opportunities of many plant viruses.

Specific alterations of viral CP subunits are achieved by targeted modifications of their nucleic acid template, *i.e.* the CP open reading frame (ORF) mediating the proteins' expression inside host or cultured production cells (see Fig. 22.2b for an example of a plasmid used for modifications of the TMV CP ORF). Site-directed mutagenesis by means of enzymes and chemical nucleic acid synthesis is applied to exchange, insert or remove nucleic acid codons for amino acids of choice. A profound knowledge on the viral shell structure (see Chap. 2) and molecular determinants of assembly and stability (see Chaps. 10, 11 and 13) is advantageous to delimit 'permissive' domains of the targeted CP sequence, allowing an insertion or attachment of foreign residues at useful sites without abolishing the viral assembly competence. Furthermore, changes in the overall isoelectric point or in locally important side chains may affect or impede capsid formation as well (in some cases involving indirect effects, *e.g.* by abolishing post-translational CP modifications promoting structural integrity, or by provoking defense reactions of the plant host inducing death of infected cells). In summary, specific alterations of viral shells are possible in many cases, but major changes of charge and structure as well as extended insertions can go along with a risk of failure and may need a series

of constructs to be tested in parallel. Successfully modified capsids, however, may then efficiently and stably accumulate in bacterial, yeast or plant cultures, yielding countless nanoparticles for numerous purposes.

22.2.3 Engineered Functionalities

Point mutations and engineered insertions of individual amino acid moieties in viral CPs typically serve as target sites for the post-assembly linkage of diverse functions to capsid scaffolds, or the localized materials deposition by distinct chemical and biochemical approaches. Peptides of up to 15 amino acids and certain relatively short protein portions often serve as anchor sequences for the interconnection with further molecules; they can be fused to ends of CP monomers, which are freely accessible in- or outside the assembled particles, or they may be integrated into surface-exposed protein domains not suffering from structural alterations. Such capture and anchoring sequences are called 'bioaffinity tags' if they mediate non-covalent specific binding of small or complex ligands, including molecules designed to serve as adaptor or linker to another layer of distinct compounds or nanostructures. Typical examples are glutathione-S-transferase portions (binding glutathione), biotin-mimicking peptides (binding streptavidin), or tetra- up to hexahistidine tags (complexing Ni ions which bridge to other molecules). Though bioaffinity tags are widely applied and in many cases efficient anchors, their non-covalent linkage to the target and sometimes poor specificity can be problematic. Alternatives lie in targets ('haptens') of commercially available antibodies which may therefore be employed as 'glue' between viral shells and any other structure or surface fashioned with the respective antibody species. Finally, advanced multi-step protein interconnection strategies lead to site-directed covalent coupling between either two proteins or a protein and a non-protein partner. They involve specific enzymatic transfer reactions addressing chemical or genetic modifications of the target protein, but have not been tested extensively with viral derivatives yet. Due to the advantageous selectivity, biocompatibility and stability of the respective coupling procedures it may be anticipated, though, that they will find their way into virus-centered nanotechnology soon.

In addition to moieties mediating an attachment between viral and technical support or cargo components, respectively, protein portions with specific biological or reporter activities are among the most sought-after constituents of nanovirologists' toolboxes. Viral shells equipped with single or multiple types of functions are considered promising high surface-area carriers for sensing and delivery purposes in both technical and medical contexts. Proteins can realize major contributions to all these applications since they may serve as detector, transducer and effector molecules, depending on their origin and, optionally, subsequent optimization. Proteins can function as environmentally sensitive fluorescent dyes; catalysts for a nearly unlimited number of biochemical transformations; highly selective capture structures for diverse targets with recognition abilities even for unique

oligonucleotide or oligosaccharide sequences; they can induce light emission, deposition of mineral crystals, undergo active motion or modulate the viscosity of natural and technical materials. These and countless further capacities of proteins may confer novel abilities to technically useful nanoparticle preparations and nanostructured 'intelligent' or 'smart' materials. In conclusion, gene technological modification of viral templates is useful for fusing proteins of interest or active domains to viral CP subunits, and will increasingly be used.

22.2.4 Specialities Worth Mentioning: Towards Synthetic Biology and Biosafety Approaches

Among the distinct goals pursued by researchers in the field of virus-based nanotechnology, some deserve special attention since they either have yielded routinely applicable methods already, or are questioning widely accepted limits of feasibility. An exceptionally well-established and powerful technique exploiting viral scaffolds is the method of phage display. Initially described in 1985, the exposure of foreign peptides or protein fragments on pre-determined domains of filamentous bacteriophages (see also Sect. 22.2.2) has been developed into versatile selection systems for amino acid sequences with desired, adapted or even novel characteristics. *In vivo* or *in vitro* mutagenesis changes the genetic information underlying oligo- or polypeptide stretches linked to the surface of filamentous phage particles. When the mutations are – in certain limits – stochastically applied, random peptide sequences are generated, which alter the population of infectious phage units within a bacterial production culture. Alternatively, a library of distinct peptide-encoding sequences may be inserted into a starting batch of phages multiplying in a bacterial cell culture. The resulting viral nanoparticles are then harvested from the culture medium and tested for properties of interest, conferred by the physically linked foreign amino acids. Favorable phages selected *e.g.* by their binding to certain substrates are extracted from the remaining population (called bio-panning) and used to start a successive round of selection. Serial selection stages will enrich most suited phages, here with high binding affinity to the substrate of choice, and thus reveal an optimized genetic information for a peptide with the desired trait.

Numerous variations of this basic method have been developed and the identified peptides tested for applications *e.g.* in the fields of enzyme design, aptamer-mediated external gene regulation, and novel affinity reagents. One can not only exploit the selection capacity of phage-based systems, but also its filamentous backbone structures for the construction of novel process materials for industrial applications. Phage display may therefore be regarded one of the pacesetters of 'synthetic biology' approaches. Another strategy in synthetic biology also creates artificial virus-like protein assemblies: Novel protein shell structures were designed on the basis of known interaction domains of plant virus CPs, exhibiting

altered subunit arrangements, sizes and/or reduced structural complexity. Such 'nanocontainers' can either assemble from single or multiple monomeric building blocks directly inside the protein expression host transformed with corresponding genetic constructs, or may be fabricated *in vitro* from purified protein species. Finally, it is worth mentioning that plants have also been exploited for the heterologous production of empty and genetically tailored, non-infectious particles of human viruses, thus excluding any risk of animal cell-derived contaminants to meet strict biosafety regulations. Those particles are tested for applications mainly in the fields of diagnostics and therapeutics, with approaches comparable to those described for plant and bacterial viruses before.

22.3 Modification of Viruses with Functional Material

The driving force to create new and more complex nanostructures is on the one hand technology, *e.g.* scaling-down of computing, memory storage, and sensor devices; in other words, more functional elements can be placed in the same volume, as exemplified by the ever decreasing size of transistors and interconnects in computer chips, and by magnetic recording bits, but also by many biosensors. The usefulness of viruses relies on their complex and well tunable surface chemistry (see Sect. 22.2).

A typical application example for a functional solid material [15] is the design of superparamagnetic nanoparticles: This requires ferromagnetic matter cut to such fine bits that the magnetic moment can be switched by thermal activation, *i.e.* it fluctuates spatially. It is as intriguing as useful that typical sizes of virions are just above this superparamagnetic limit (a few nm up to about 20 nm size). Such particles, suspended in liquids, can be directed by an external magnetic field focus, which is useful for marking tumors. On the other hand, an oscillating magnetic field (with a flux of the order of 0.01 T, and a frequency on the order of 100 kHz) induces highly localized heating, which can destroy tumors. It is obvious that the exact size, shape, and chemistry of the particles is crucial. Spherical virus shells, similar to apoferritin, offer a much higher definition than particles synthesized by standard methods, and excellent biocompatibility.

As in this example, most functional materials require metal compounds, and so the construction of the nanoobjects is based on standard inorganic chemistry of metal ions, such as precipitation and redox reactions (Table 22.1). One interesting aspect is the interface to the virus: First, the usual metal ligands are incorporated in the virus surface. Here amine and thiol groups, often genetically engineered, are most popular, but also hydroxy groups can work very well. Second, a future nanotechnology (in a strict definition) could be built very elegantly on a combination of biology with chemistry and nanoscale physics [16]. The key is the combination of biochemical synthesis methods with inorganic methods. The advantage of employing virions or viral capsids as scaffolds is on the one hand a better definition of the nanostructures, compared to template-free syntheses; on the other hand,

Table 22.1 Examples for chemical deposition reactions of functional materials on viruses in aqueous suspension

$Au(III)_{ads} + 1.5\ BH_3 + 1.5\ H_2O$	→	$Au + 1.5\ BH_2(OH) + 3\ H^+$ (ads. and chem. reduction)
$Ag(I) + e^- + h\nu$	→	Ag (photoreduction)
$Ni(II) + BH_3 + H_2O$	→	$Ni + BH_2(OH) + 2\ H^+$ (electroless deposition, catalyzed by Pd and Ni)
$Si(OC_2H_5)_4 + 2\ H_2O$	→	$SiO_2 + 4\ C_2H_5OH$ (hydrolysis, sol-gel procedure)
$Ti(CH_3)_4 + 2\ H_2O$	→	$TiO_2 + 4\ CH_4$ (repeated; atomic layer deposition)
$Zn(II) + 2\ OH^-$	→	$ZnO + H_2O$ (catalyzed by Pd in presence of nitrate)

principally superior methods such as synthesis from small precursors (*e.g.* from polynuclear metal complexes) or even by manipulation of each single atom are far from being practical. A big challenge is developing recipes and especially general rules, such that known inorganic material synthesis methods can be tranferred to virus interfaces.

Organic coatings are quite unusual since the biochemical functionalization of the viral surfaces can be used to obtain a huge variety of chemical groups. However, in some cases simple or fast methods are searched for, such as coating with a polymer. Although the van der Waals forces between a polymer chain and a virus are huge, binding polymers is based on short-range chemical forces, such as hydrogen bonds. Not only a polymer synthesis, but any other synthesis at viral interfaces, requires careful balance between van der Waals, chemical, and electrostatic forces [17] (solvation *e.g.*) – exactly the same "colloid chemistry" that is required to bind nanoparticles to viruses, or viruses to viruses. In fact, some virus systems (TMV and fd phage) have contributed themselves significantly to the development of colloid chemistry since they offered the required submicro size, and defined surface charges.

Turning to the future, the highly complex chemistry of the virion surface allows building nanodevices with various, even multiple functions and physical properties. However, multifunctional nanoparticles are quite rare, and very few examples are known for virus templates. The decisive advantage of viruses would be their integration into micro- or macroscale solid devices: Here biochemical reactions on the solid surface, with their huge specificity, could make this process automatic (highly parallel "self-assembly", "bottom-up"). Possible solutions could be based on binding antibodies or parts of the virus (nucleic acid, capsid protein) to the surface. Such a strategy would work on the molecular scale, and has thus the potential to revolutionize nanofabrication. The standard nanofabrication is based on micro- or even nanoscale placement of the functional material, by lithography ("top-down" structuring – a widespread method is electron beam lithography, where the beam modifies a polymer layer pixel by pixel, see Sect. 22.5.3). This means that (nanoscale) devices have to be built sequentially, *i.e.* slow, while molecular recognition and self-assembly build complex molecules and nanostructures very quickly from smaller units, which assemble in a well-defined way, and on a smaller scale. The difficulty is to have a functional coating and also a biochemical function – the chemical groups should not interfere much with the function, and vice versa the coating must not obstruct the chemical groups.

22.4 Hierarchical Assembly into Complex Structures

Hierarchical self-assembly is characteristic of biological systems. It can be described as *formation of an ordered structure through a set of self-assembly steps*, *which decreases in strength* [18]. In other words, the term *hierarchical* refers to the situation where each self-assembly step is a guide for the next one. One of the better-known and understood examples of hierarchical self-assembly is the formation of TMV. Upon certain conditions, single units of a capsid protein interact with each other to form larger aggregates and discs (Fig. 22.3a). Assembly of a TMV is then initiated by a binding of an RNA origin of assembly loop to a two-layered disc (20S) comprised of 34 copies of a CP (Fig. 22.3b). Subsequent particle formation involves transformation of the 20S disc into a short helix and concomitant incorporation of the RNA strand between the CP layers. The resulting nucleoprotein helix is elongated bidirectionally *via* stepwise addition of further discs to one end of the growing nanotube, while the other end of the virus is completed by help of smaller CP oligo- and monomers. Such strategy of building up a capsid protein shell step by step around the nucleic acid has been observed in many other viruses with different geometries, namely polyhedra (see Chap. 12). More complex viruses often require much more complex tools, such as molecular motors, to pack the genetic material (see Chap. 12). Obviously, a very advanced nanotechnology would also make use of such nanoscale machines; here the focus is on rather simple cases, which can mainly be addressed as "self-assembly". It is important to note that in many cases the hierarchical assembly process is governed (at least in its initial stage) only by the information encoded in the protein itself, with the pH as the self-assembly triggering signal. This fascinating process, rather common in nature, is becoming an important experimental tool for the fabrication of complex mesoscopic structures, which often exhibit unique properties surpassing those of individual components.

One of the simplest methods of organizing particles into an ordered hierarchical structure is drying. By varying the virus and the salt concentration in solution, as well as the substrate surface properties, it is possible to tune the pattern formation during drying (see Fig. 22.3). Various structures can form in a capillary tube and on planar surfaces [19]. Stripe patterns or continuous films of filamentous viruses, of different thickness and orientation, are relatively easy to obtain with this technique from filamentous viruses. The ability to control surface arrangement of biomolecules promises new opportunities in sensing technologies or tissue engineering [20].

Another possible way in which viruses can form a hierarchical assembly is end-to-end interaction. In the case of TMV, the length of an individual virus is determined by the length of the RNA, but complete virions can arrange into a longer one-dimensional (1D) structure *via* interaction between proteins exposed at both ends of the viral rod, creating stacks of viruses (see Fig. 22.3). This ability is of utmost technological importance. 1D structures like nanowires or nanotubes are key components in a broad range of nano-applications. One interesting example is the synthesis of long, conductive polymeric composites of filamentous viruses, comparable to metal-coated structures discussed in Sect. 22.6. Such structures can have very high aspect ratios, and will probably find use in electronics, optics, sensing, as well as in biomedical engineering.

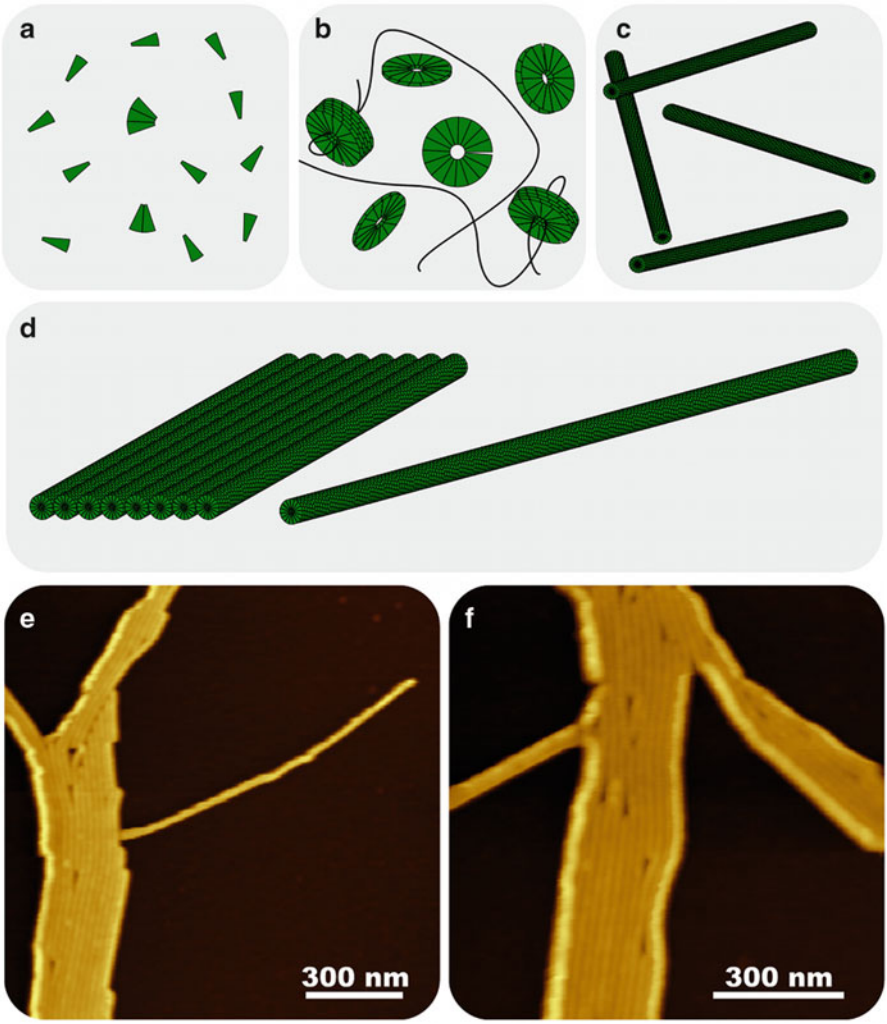


Fig. 22.3 Assembly processes of TMV. CP subunits form small aggregates (**a**), which assemble to form oligomers, rings and ring-stacks (**b**). An RNA origin of assembly positions itself within a two-layered 20S CP aggregate, which is transformed into a nucleoprotein helix, and recruits further rings and oligomers until the whole RNA is buried within the protein shell (**c**) *via* bidirectional assembly. For details, refer to Fig. 22.9. 1D structures can be extended by linear assembly when virus fibers attach to each other by end-to-end alignment, (**d**) *right*. Two-dimensional structures are analogous to nematic liquid crystals, but often imperfect, (**d**) *left*. Experimental examples (AFM) for TMV on a surface (**e**), (**f**) (image from M. Gorzny and S.D. Evans)

Rational assembly strategies for complex hybrid structures are quite rare: They can be founded on chemical and also on genetical modifications of CP surfaces, for which a set of standard methods is available. A bacteriophage can be programmed to selectively bind a chalcogenide nanocrystal at one of its ends, and metal particles on the capsid surface. Simple mixing of the phages with both nanoparticles results in such a complex structure, and further assembly can produce various architectures from 1D

to three-dimensional (3D). An example was investigated by Lee et al. [21], who exposed genetically modified M13 to ZnS nanocrystals, which spontaneously evolved into a hybrid bioinorganic film, extending to the macroscopic range (centimeters). While such combined approaches are typical of self-assembly on the laboratory scale, there is much thrust to use them for manufacturing electronic circuits in a truly "bottom-up" way.

Self-assembly strategies, as attractive as they may appear, intrinsically work in the microscopic world, based on interactions well below micrometers. Without highly sophisticated steering mechanisms, such as the ones nature developed in cells, they will not be of advantage for producing macroscale objects. Tackling the increasingly complex architecture of the nanoscale world may be much more rewarding. Typical examples are core-shell structures, for which ultrasmall chunks of organic or inorganic materials are coated by thin layers. Viral capsids can play the role of such layers, and since empty cavities are intrinsic properties of many plant viruses, the reverse approach is possible, and a range of mainly spherical virus shells have been filled with especially metal oxides and metals (see Sect. 22.3). This feature is of tremendous importance for many applications, out of which the most promising are drug delivery and guided growth of inorganic nanostructures like nanowires or nanoparticles.

When the viral shell is built up step by step around a given particle, the method is again a coating strategy. However, viruses allow for a much higher precision than chemical reactions or adsorption [22]: With the trick of binding the RNA assembly origin (RNA packaging signal, see Chap. 12) to a gold nanoparticle, simple mixing with CP subunits can result in assembly of virus-like particles around the gold nanoparticle. This artificial structure has the morphology and the chemical features characteristic of a virus. This strategy of employing the CP self-assembly processes to envelope an inorganic cargo could be used in many applications, *i.e.* drug delivery (see Sect. 22.8) or catalysis. The morphology and the chemical composition are thus more precisely tuned than possible with organic layers.

22.5 Nanoscale Analysis and Manipulation

Physics provides nowadays techniques that allow direct visualization, physical analysis and manipulation of single nano-objects including individual molecules. These techniques include atomic force microscopy (AFM; described in detail in Chaps. 8 and 18) and optical and magnetic tweezers (described in detail in Chap. 9). Here only a reminder of a few aspects of these techniques as applied to viral nanotechnology is provided. In addition, other physical techniques that are being used in viral nanotechnology are briefly described in this Section.

22.5.1 Scanning Probe Techniques

Scanning probe microscopy (SPM) techniques were originally developed for hard surfaces. Among these techniques, AFM (see Chap. 8 for a detailed description) has

proved to be an efficient tool to answer many fundamental questions concerning soft matter, including animal viruses, plant viruses and bacteriophages (see Chaps. 8 and 14). With AFM, viruses can be imaged in physiological media. Structural and morphological changes of the virus can be monitored in response to a change of the environmental conditions. Hence, this technique is a complementary tool for the immediate identification of viruses. Most important are contact and noncontact (mainly "tapping") AFM, with a vertical resolution of below 1 nm, although lateral resolution values below 10 nm are hard to reach. A large virus particle means a large vertical tip movement, which is usually not well compatible with high lateral resolution. However, special techniques allow to obtain more information: AFM can scan the interior architecture of viral particles, after systematically stripping away layers of their structure (see Chaps. 8 and 18). Plant viruses in crystalline form as well as single viral particles and bacteriophages can be visualized by this technique. Contact AFM is normally not suitable for scanning, but it can provide information about a variety of mechanical properties of individual viruses, including local adhesion and elasticity of capsids, although other methods like tapping or "jumping" modes (see Chap. 8) are generally more suitable. The Young's modulus and the Poisson ratio can be determined from nanoindentation experiments performed with AFM [23–25] (see Chap. 18).

In contrast, scanning tunneling microscopy (STM) is only useful for conductive and rather flat surfaces since it uses a conductive tip in close proximity (<1 nm), which however is not defined in shape (with the possible exception of some atoms at the tip apex). On poorly conductive objects such as viruses, the tip penetrates the object completely in order to reach the underlying surface. Even when the virus is coated by a thin metal layer, the relatively large height differences of up to tens of nm makes imaging difficult. Still, some reliable imaging conditions were found, most remarkably for usually nonconductive substrates such as glass, which acquire some conductivity upon hydration [26]. STM is also useful for detailed highly local current-voltage (I/V) measurements, which are however restricted to composites of viruses with metal or semiconductor coatings.

A further evolution of SPM is the dip-pen nanolithography. This technique uses an AFM tip as a nib, a solid substrate as paper and molecules with chemical affinity for the solid substrate as ink [27]. Plant virus and phage nanostructures can be fabricated by capillary transport of linker molecules from the AFM tip to a flat solid substrate, followed by incubation with the virus suspension. The linker molecules have to be bifunctional: One moiety binds to the surface, arranged by the tip in the desired micro- or submicro pattern, and the second one is employed for chemoselective attachment of the viruses. The control can reach the single particle level, when the pattern size and spacing is chosen appropriately.

22.5.2 Optical Tweezers

Optical tweezers (see Chap. 9) use a highly focused laser beam to provide an attractive or repulsive force to physically hold and move microscopic dielectric

objects. The basic principle behind optical tweezers is the momentum transfer associated with bending light. If an object bends the light, changing its momentum, the object undergoes an equal and opposite momentum change. This gives origin to a force acting on the object. In a typical setup, the incoming light comes from a laser which has a Gaussian intensity profile. When the light interacts with an object, the light rays are bent according to the law of refraction and reflection. The forces from such rays can be split into two components: the scattering force pointing in the direction of incident light and the gradient force coming from the gradient of the Gaussian intensity profile and pointing towards the center of the beam. The gradient force is a restoring force that pulls the object towards the center. If the contribution to scattering force of the refracted rays is larger than that of the reflected rays, then a restoring force is also created along the beam axis, and a stable trap will build up. Individual virions and oriented arrays of virions can be optically confined within volumes of a few cubic micrometers without damage, and manipulated in space with a precision of about 500 nm [28]. The very low and precisely tunable forces can be used to probe subviral structures, such as the packaging of DNA in phages [29] (see Chaps. 9 and 12).

22.5.3 Electron Beam Lithography and Photolithography

Nanofabrication techniques, such as electron beam lithography (eBL), permit the handling of single virus particles. eBL is a maskless lithography technique that uses a focused beam of electrons to pattern substrate surfaces covered with a polymeric film resist (tens to hundreds of nm thick). This technique is the simplest way to produce self-designed structures below 50 nm lateral feature size (even $<$10 nm can be reached). After electron bombardment, exposed or non-exposed regions of the resist are selectively removed in a developer solution. Negative resists form bonds or cross-links between polymer chains during e-beam exposure and turn out insoluble when developed, creating small areas of protruding structures. On the other hand, positive resists undergo bond breaking when irradiated, as a result exposed areas become more soluble in the developer. Therefore, small areas of recesses are formed. Depending on the desired design, positive or negative resist nanostructures are thus used to transfer material to the substrate, either by deposition of metals, or by etching. This technique has been widely applied for prototype integrated circuits, and for nanotechnology architectures. TMV is compatible with electron beam lithography processes and can be integrated in nanostructures made of positive and also of negative eBL tone resists. Viral particles maintain their biochemical functionality after fabrication steps, which was verified through selective immunocoating of the TMV [30] (see Fig. 22.4).

Photolithography is of much wider technical use. Here the structuring beam is ultraviolet (UV) light. Patterns are not formed by scanning, but by exposing a

polymeric film resist (the substrate) to UV light through a photomask, in one single fabrication step. The photomask is an opaque plate with holes (or windows) that allow light to pass through in a defined pattern. Such structures are the basis for microchip mass fabrication. The most advanced structures reach the deep nanoscale (<50 nm), but they require extreme UV and even soft X-ray optics, which are very rarely used in research laboratories. However, these methods are very suitable for arranging viruses, and as yet practically not explored.

22.5.4 Electric and Magnetic Fields

Manipulation of plant viruses can be based on the application of electric and magnetic fields. Best known is electrophoresis in porous media (gels) for particle separation; however, this is limited to the macro- and microscale. More suitable is dielectrophoresis (DEP), where viral particles move induced by the polarization effects in a non-uniform electric field (see Fig. 22.5). The non-uniformity (the field gradient) is induced by microstructured electrodes, and can thus work on the nanoscale. Indeed, DEP is known generally as a method to orient and to move colloidal particles, hence the application on viruses is straightforward. Movement is caused by the unbalanced force of the non-uniform electric field on the viral particle's induced dipole moment: one "end" of the dipole is in a weaker field than the other, causing the particle to be pulled electrostatically along the electric field gradient. Plant viruses as TMV and CPMV can be accumulated and oriented at microstructured electrodes using this technique [31, 32]. TMV presents a high polarizability, since it lacks an insulating membrane. Therefore TMV exhibits positive dielectric behavior (it is attracted to the electrodes) or negative (it is repelled from electrodes) depending on the frequency of the applied electric field. Likewise, CPMV presents positive dielectric behavior at low frequency of the applied electric field (2–3 MHz), and negative dielectric behavior at higher frequencies (18–20 MHz) [31].

In analogy to the electrical polarizability, also the magnetic susceptibility can be highly anisotropic. However, since this effect is based on the (very weak) diamagnetic moment, which is present in all matter, a rather large flux (1 T range) is required for orientation, and a large field gradient for movement. The orientation effect is well known from the alignment of nematic liquid crystalline and colloidal crystalline samples, especially for TMV and fd, which are staples in the research of liquid crystals [33]. Small-angle X-ray scattering is a typical detection technique. Fields and field gradients can be combined and varied in time to achieve a right-circular or left-circular orientation of TMV [34].

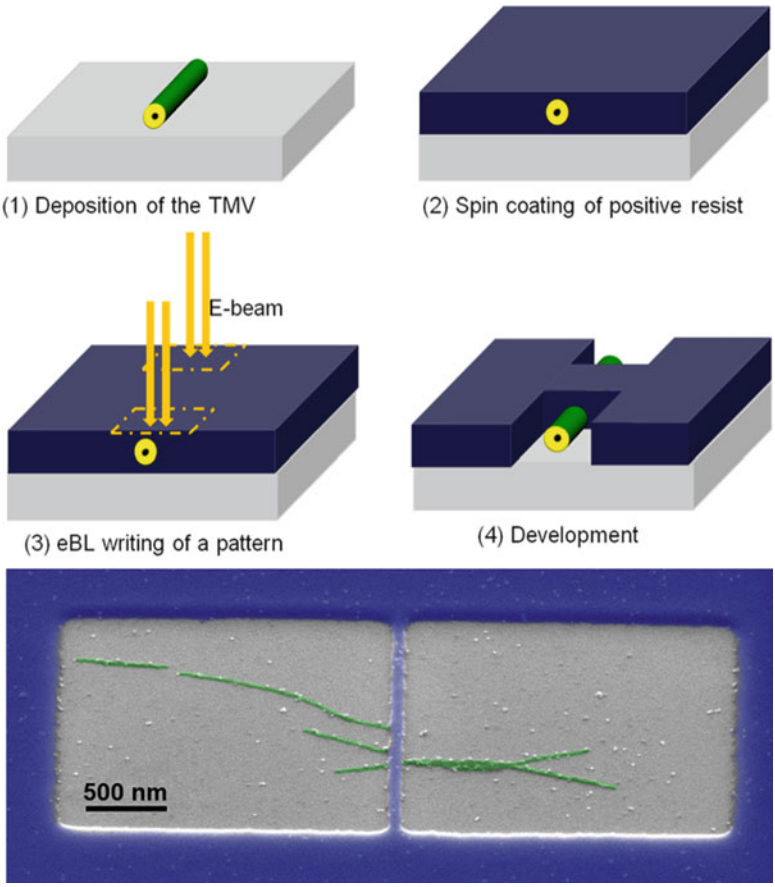


Fig. 22.4 Integration of TMV into typical electron beam lithography production steps. Typical nanofabrication schemes involve polymer coating, electron beam structuring, and selective dissolution of the irradiated polymer. *The bottom image* is a scanning electron micrograph that shows linearly assembled virus particles below a segmented rectangular polymer window produced in this way [30]

22.6 Viruses as Templates

A typical strategy in nanoscale science and technology is to employ a nanoscale object to achieve a certain function. The object – here a virus – provides proper size and shape, and a coating or filling bestows the function. The material, mainly inorganic, is difficult to tailor – that is why a template is required. In such strategy the virus template plays no biological role, so one could call it "inactive". The material on or in virions would then be "active", and tailored on the nanoscale. It might exhibit physical properties that are different from those of the bulk material, such as quantization of electrical conductance, (dis)appearance of

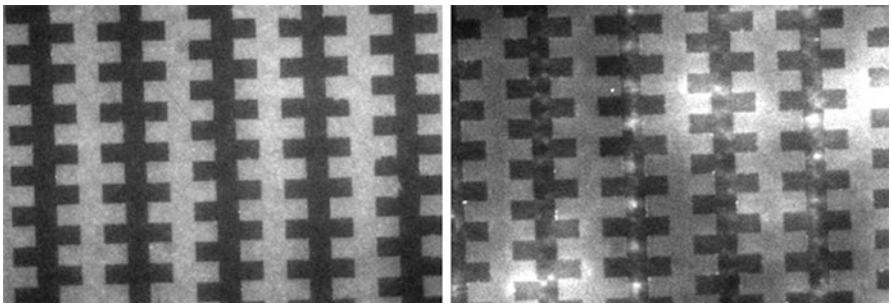


Fig. 22.5 Dielectrophoretic enrichment of plant viruses at microelectrodes. Switching on an alternating electrical field (20 V, 1 MHz) leads to the appearance of fluorescence (*white spots*) due to high local CPMV concentration. The frequency-dependent polarizability of the virus is at the origin of this phenomenon [31] (image from I. Ermolina)

ferromagnetic coupling, changes in light absorption and emission, or simply a high ratio surface atoms/bulk atoms. In this respect, inorganic structures (*e.g.* metals, metal compounds and ionic crystals) and organic/inorganic composites are often more attractive than organic molecules.

While nanoparticles are produced since many decades (and in some cases known for centuries), a relatively new aspect is the controlled fabrication of devices (see Sects. 22.7 and 22.8), and integration of nanostructures – both with nanometer precision. Linear structures (wires, tubes, chains of clusters) play a major role as spatial directors and as nanowire interconnects. The advantages of employing coated viral fibers, compared to conventional nanowires, are a better spatial definition of the nanostructures, and better defined functional groups. Moreover, principally superior ideas, such as nanowires that are defined to atomic detail, are not in reach.

A good example is the anisotropic shape of material on or in tubular virions [35], which can align the magnetization direction with the virion axis. The best known examples are gold-coted virions; in fact, the history of gold nanoparticle binding dates back to before 1939, when the first images of gold on virions were recorded [36]. The motivation was the shift of the optical absorption, resulting in deepening of the red color, used for virus detection. The specificity is too low for most uses, and it is more likely that this method will be useful to design highly defined plasmonic nanostructures. A third application example concerns electrical conductivity that was shown for Pt-covered TMV [37]. However, most modifications have as yet not found any application.

22.6.1 Templates for Material in Viral Cavities

Many RNA bacteriophages are tightly packed nucleoprotein complexes, with no internal cavities. Many DNA phages, some RNA phages and most plant viruses, in

contrast, have either a spherical cavity, which is filled by RNA or DNA, or an internal tubular channel, which is empty (filled by the surrounding liquid, and in dry environments likely by condensed water). Spherical viruses can in many cases be disassembled into protein and nucleic acid, and the pure proteins can be reconstituted into the empty shell, which should now also be filled by electrolyte (see Chaps. 10 and 11). Exactly this procedure is well known from a related nonviral shell, apoferritin. Apoferritin is filled with a range of oxides and metals, based on the chemical reaction types described in Sect. 22.3, and the same strategy can be used for spherical viruses.

A successful filling by solid material will depend on this reaction, and on the chemical functionalities present on the virus. It is crucially important to avoid any reaction on the exterior surface, hence the groups, their charges, their reactivities should differ as much as possible. It turns out that these conditions cannot be easily fulfilled, even for genetically engineered viruses. Moreover, the final material size is mainly well below 5 nm – template-free strategies cannot easily reach this range, so this could develop into a real chance for the technical use of viruses. However, as yet, only few examples of materials encapsulated in viruses are known. Some spherical viruses can be disassembled, the RNA removed, and reassembled. Ions can diffuse through pores in the shell to the interior cavity, and react in analogy to the examples described in Sect. 22.3. Iron oxide particles inside such a small cage are remarkably small, so small, that magnetic couplings accumulate to unusually small moments. This superparamagnetism is a very typical difference to larger particles, and very useful for ferrofluids (see Fig. 22.6).

Many filamentous viruses have a tubular channel. The concentration of certain amino acid moieties in the wall is so high that cations and some metal complexes show a preference for it, and materials can be grown inside. The frequent observation of small particles suggests a mechanism of nucleation, followed by rapid growth. When the material fills the channel, *i.e.* when it produces a wire, liquid is encapsulated in the virion [39]. This explains why a given channel cannot contain more than two wires. This reaction is best known from two-step electroless metallization. The yield and control are probably not yet sufficient to merit production of useful amounts of wires. Optical, electrical, and magnetic properties of the wires should differ hugely from bulk material, the main reason being the spatial confinement of the electrons.

22.6.2 Templates for Material on Viral Surfaces

In principle, the argument for material synthesis in a cavity can be reversed to achieve full selectivity for the external capsid. However, in most cases such thick layers of material are produced that a potential presence of it in a cavity has no influence. The exterior surface of virions is easily accessible, so a plethora of organic reactions (Sect. 22.2), adsorption, and of solid material synthesis are available (see Fig. 22.6). This is closely related to surface chemistry and surface

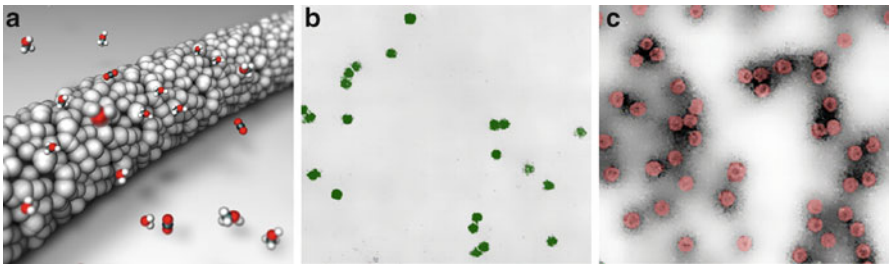


Fig. 22.6 Coating and encapsulation of inorganic materials. (**a**) Model of TMV coated by a layer of platinum nanoparticles. This structure can be used as electrode for fuel cells that oxidize organic compounds to carbon dioxide and water. (**b**), (**c**) Encapsulation of iron oxide in CCMV. Without staining, TEM shows only the high density core (**b**), with selective staining the whole particles is visualized (**c**) (reproduced from [38] with permission)

science, with the surface here being the virion. However, many methods (see Sect. 22.3) produce rough and thick (>50 nm) coatings, due to badly controlled nucleation or growth rate. Such structures can often be produced in simpler ways, hence growing materials on viruses might not be competitive. In some cases however, mainly for tube- or wire-like phages and viruses, the result is still useful because other ways of synthesis are complex. Magnetic coatings turn the template into a nanoscale magnet (Fig. 22.7), with uses in ferrofluids for advanced vibration damping. Ferrofluids are based on a much increased viscosity whenever a magnetic field is switched on; they require nanoscale magnets.

When viruses are assembled in parallel fashion, metallisation can produce a rather well-defined surface of high porosity. Based on this idea, various metals and oxides on M13 phage and on TMV have been tested as electrodes in nickel and in lithium ion batteries [41, 42] (see Fig. 22.8). Improved control of roughness and of parallel assembly can now open the way to use virions in electronic devices (Sect. 22.7).

22.6.3 Double Templating

Such systems have to be based on those discussed in Sect. 22.6.1. An additional reaction step, based on that described in Sect. 22.6.2 will now produce sphere-in-shell systems, or coaxial wire-tube systems. These can be very attractive since there are very few other methods to design matter in 3D so finely. For example, ferromagnetic cores are coated with antiferromagnetic shells to produce a coupling that can lead to exchange bias (asymmetric magnetic hysteresis curve) – this could be done much simpler with a spherical virus that is first filled in the core. There is a huge drive to applications such as ultrahigh density magnetic storage, where each particle would make up one magnetic bit. However, the required fabrication steps are as yet not known, and analysis methods on the required scale, *i.e.* <5 nm for the interior structure, are scarce. Indeed such small structures lead us to the limit of today's nanotechnology.

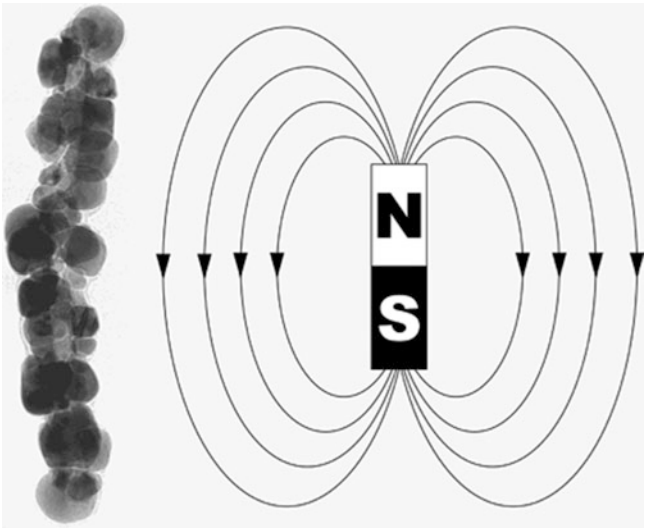


Fig. 22.7 Cobalt-plated plant virus as nanoscale bar magnet. Plant viruses, here two TMVs, can be coated by first adsorbing noble metal complexes (catalytic precursors), then reducing them and applying a metallization solution. Here cobalt coating was achieved [40]; such rather thick layers are ferromagnetic, hence a filamentous virus transforms into a nanoscale bar magnet (reproduced from [40] with permission)

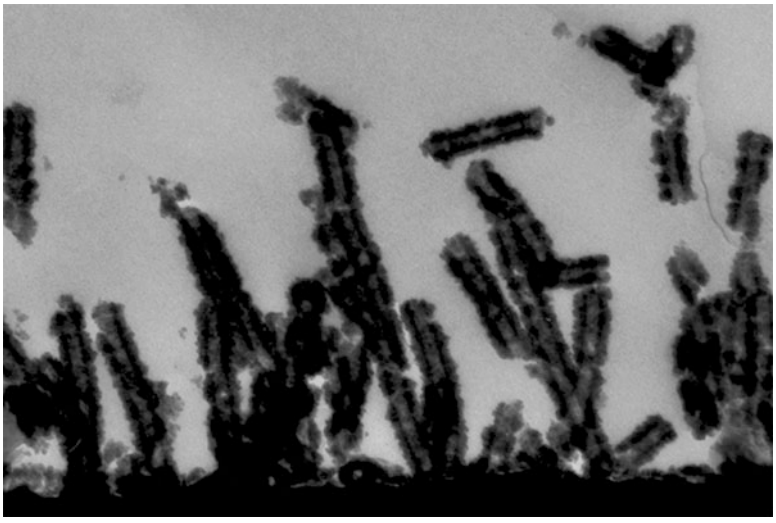


Fig. 22.8 Battery electrode made of nickel-plated TMV. The virus has to assemble standing-up on the flat substrate (*black*, at *bottom*), which is achieved by a Cys mutation. The virus layer is then treated with Pd(II) as catalytic precursor, followed by electroless deposition of a nickel layer. In this way, a highly porous surface is obtained [41] (TEM image from J.N. Culver)

22.7 Electronic Devices

Nanostructures made by using viruses as templates have been synthesized for over 10 years. Despite intensive research, most of the published work has concentrated on the synthesis. Indeed there have been only a few reports showing technological applications of virus-inorganic nanostructures. One can distinguish applications that require mass fabrication, *e.g.* ferrofluids and energy generation or storage, and those which require relatively small amounts, *e.g.* biosensors or nanowires (electrical connectors). The following paragraph we demonstrates these concepts with a few examples.

A good example of taking advantage of the virus morphology has been demonstrated by Royston et al. [41]: Engineered TMV assemblies, with cystein residues incorporated into the capsid protein, were vertically patterned onto gold substrates *via* gold-thiol (cystein) interaction. Nickel and cobalt were then deposited by electroless deposition. This layer of vertically assembled metal-coated viruses was subsequently used as an electrode in a nickel-zinc battery (Fig. 22.8). Tests showed that the incorporation of virus into the electrode increases its surface area and in consequence doubles the total electrode capacity. In a similar spirit, platinum-coated TMV has been tested as an anode material for a direct methanol fuel cell [37]. For this, TMV has been decorated with fine platinum nanoparticles by means of electroless deposition. During the growth, Pt nanoparticles interconnected with each other, thus creating a continuous and thin metallic shell around the virus (Fig. 22.6). Such hybrid Pt-TMV nanotubes formed a so called support-less catalyst, which in tests overperformed (higher surface area and stability) conventional, nanoparticle-based catalysts. Utilization of TMV-based catalysts is hence a prospect for reducing the amount of Pt (costs), and also for improving the cell efficiency. For microelectronic devices, viruses have not yet been used extensively. Some of the most impressive results were obtained with apoferritin, *e.g.* for floating memory [43]. Metallized virus-based nanostructures, especially 1D-like nanowires or nanotubes, are natural candidates for new types of electrical interconnectors. Electrical properties of virus-based nanowires/nanotubes were investigated by rather few researchers, but it emanates that conductive structures based on viruses can be exploited [44]. Metal/virus composites embedded in polymer films may even show memory effects [45].

Due to the size and linear morphology TMV has been found to have a profound effect on magnetoviscosity of a cobalt ferrite-based ferrofluid [46]. Simply by mixing a commercial ferrofluid with a TMV suspension one can observe the increase of magnetoviscosity. Moreover, the magnetoviscosity has been found to be less susceptible to a shear thinning (*i.e.*, a reduction of the viscosity with the applied stress) which is a desirable from the technological point of view. This new class of compound could potentially be used in micromechanical dampers.

Since viruses are nearly exclusively used as (advanced) scaffolds for active elements (nanoparticles or metallic layers), the design of a functional device based solely on a biological structure is especially challenging. An excellent example has been demonstrated by a Lee et al. [47]: Genetically engineered M13

phages with overexpressed pVIII N-termini (with a variable number of the negatively charged amino-acid glutamate) were allowed to assemble on the surface (see Sect. 22.4) of a gold film. Putting another gold electrode on the top creates a Au-M13-Au sandwich. Such structures exhibit a piezoelectric effect capable of producing up to 6 nA of current and 400 mV of potential capable to power up a small liquid crystal display.

22.8 Biochemical Detection Arrays and Targeted Drug Delivery

Besides their amenability to chemical modifications and materials synthesis or encapsulation, which typically result in bio/inorganic hybrid structures or conjugates with small organic compounds, the protein surfaces of viral templates offer unique opportunities for the fabrication of biochemically active 'smart' materials and arrangements with complex activities. As outlined in Sect. 22.2, selective capture or attachment as well as signal transduction capacities are amongst the technically most attractive functions of proteins. Their effective utilization in sensing or targeting devices, however, still poses a number of challenges. (1) First of all, the active sites of proteinaceous effectors need to be freely accessible to the target compounds. Furthermore, full reactivity often demands for conformational flexibility of the polypeptide chain. Both are best achieved by an exposed position of the functional unit on a suitable nanostructured scaffold in a controlled orientation. (2) Second, non-specific binding of the target molecules to the scaffold surface should be avoided, for which a different type of protein may be an ideal 'blocking' compound. A close vicinity of other proteins to the reactive one was also proven to generally promote and preserve biochemical activities. (3) Third, a high, but adjustable density of the active units on a polyvalent carrier structure with multiple immobilization sites is essential to ensure optimum functionality in different environments. (4) Finally, an efficient and targeted integration of the carrier into a device or a biological environment should be possible. Current protein-employing sensing arrays or targeting nanocapsules, however, usually cannot fulfill all of these demands and thus suffer from several limitations. While *e.g.* sterical hindrance can be reduced by fixing active protein units on technically fabricated nanopillars or networks of synthetic phases, this strategy alone is not sufficient to counteract adverse surface effects on both targets and active protein sites. An attachment of flexible linkers such as nucleic acid fragments or bioaffinity tags to pre-selected amino acid stretches of the bioactive unit can improve its reactivity due to controlled orientation and spacing to the support, but still needs additional measures against unwanted interactions with it. The immobilization reaction itself relies on the presence of suitable addressable groups on the carrier surface, which are lacking on typical array materials and thus are generated *e.g.* by additional coating layers. These should supply predictable reactivity for stable interconnections without negatively affecting protein or reactant integrity. Though numerous formulations of such 'adhesives' have been developed for the one- or two-step

fixation of proteins, their binding efficiencies and side reactions are still difficult to control in many cases.

Compared to assemblies on conventional planar, or advanced nanostructured synthetic substrates, peptide or protein ensembles exposed on multivalent viral capsid surfaces therefore have numerous advantages. Realized by means of biochemical linkage to pre-defined target sites, or by genetic fusion to selected amino acid portions of all or a subset of the viral CP subunits, they meet indeed all the requirements for highly active versatile arrangements listed above. In conclusion, virus-scaffolded protein arrays are most favorable and technically promising composite materials, applicable for both sensing and targeting purposes as explained in the following.

22.8.1 Display of Capture and Targeting Functions on Viral Shells: State-of-the-Art

Protein or peptide portions with selective affinities, presented on the surface of viruses or VLPs, offer numerous possibilities for their use in detection systems. Most compatible with existing read-out technology is an application of the viral templates as polyvalent adaptor phase between a technical support and the biochemically active units. The virus-derived templates thus will arrange, position and stabilize target-binding amino acid domains of one or more types in close vicinity to each other. While small peptides capturing or complexing ions, metals or simple compounds have been directly exposed on virus capsids by means of genetic fusion (predominantly on bacteriophages), larger functions catching compounds with *e.g.* medical or environmental relevance are mostly attached by chemical linkage or standard bioaffinity tags as described in Sect. 22.2. In this context, the typically employed proteins are antibodies or engineered derivatives thereof. Antibodies immobilized on spherical plant viruses such as CCMV or CPMV react for example with pathogenic bacteria or surface markers of distinct cell types (for these and numerous further references see [10] and [48]). In principle, also a direct *in planta* production of VLPs coated with extended capture domains or other additional protein fragments is possible, given that the genetically tailored fusion proteins assemble despite the foreign portions. Careful design of both the linking strategies between the viral CP and the non-viral domain, and molecular tools optimizing the production kinetics of the respective constructs have allowed to harvest functionalized chimeric nanoparticles from plant expression hosts at high yields. Stiff tobamovirus (turnip vein clearing virus TVCV) rods displaying antibody-binding fragments of *Staphylococcus aureus* protein A [49], flexuous potato virus X particles covalently coated with antibody derivatives directed against a herbicide [50] or spherical empty CPMV virus-like shells presenting GFP [51] may exemplify this most straightforward and thus economically promising strategy, which,

however, is less universally applicable than *ex-situ* linkage between viral shell structures and functional moieties up to now.

22.8.2 Perspectives of Virus-Scaffolded Detection Arrays

The utilization of accordingly functionalized virus-like nanoparticles in detection systems then might follow distinct approaches. Numerous protocols for the deposition of viral capsids on various useful substrates are available, controlling not only surface coverage, but also bonding strength, precise position and even orientation of the nanoparticles (refer to Sect. 22.4 for additional information). Advanced procedures allow for the fabrication of patterned and layered three-dimensional arrangements with blends of distinct and, where appropriate, interacting functionalities. These may include also conductive metal components allowing for electronic measurements [5, 52], see Fig. 22.9. The resulting arrays are high-performance capturing units which can serve as concentrator and indicator for specifically addressed targets. They may either be employed in enzyme-involving diagnostic assays such as ELISA (enzyme-linked immunosorbent assay) formats in combination with secondary binding partners *e.g.* for a colorimetric compound detection, or they can be integrated into sensor devices exploiting electrochemical or physical read-out technology to detect captured analytes, *e.g.* by the generation of electrical currents, changes in conductivity, surface plasmon resonance or direct quantification of adsorbed material by a quartz crystal microbalance. A coherent overview on analytical principles applicable for virus-based sensing has been published by Mao et al. [55].

Finally, three additional options to further improve sensing systems by way of viral derivatives should be noted. First, certain virus templates may not only be deposited on technical supports post-assembly, but may be grown from their building blocks "bottom-up", *i.e. in situ* directly on attachment sites pre-defined by the local presence of an anchor molecule. This strategy was developed recently for TMV-derived nanorods on patterned polymer substrates (Fig. 22.9) [54] and may help to generate freshly functionalized carrier templates on demand, prior to their use even in difficult-to-access reaction chambers of nanosized future devices. Second, for increasingly diverse applications, planar detection microarrays are being replaced by 'liquid' or 'bead arrays' these days, with the selective binding reactions taking place on accordingly functionalized nanoparticles in suspension. Combinations of internal bead-specific labels (barcoding the affinity of the respective bead species) and flow separation technologies enable highly sensitive multiplexing approaches allowing simultaneous detection of distinct analytes. Since nanobead preparations can be densely equipped with viral templates as well, the resulting composite beads with their amplified soft matter surface-area and additional degrees of freedom for combinatorial activities on every functionalized bead are promising candidates for further improved liquid arrays. Third, the nanoparticulate nature of the polyvalent viral capsids themselves offers the striking opportunities to use them directly as array particles, or as signal-enhancing effectors in both conventional and novel detection

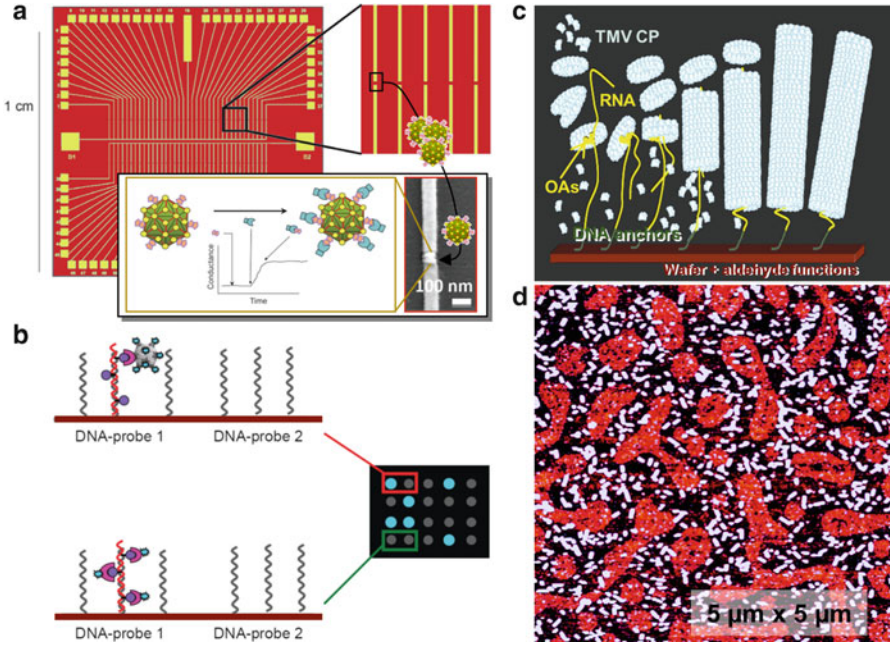


Fig. 22.9 Biosensing and positioning by help of viral particles. (**a**) Electronic sensor chip with many gold electrode pairs with common drain. Gaps contain dual-functionalized CPMV carrying *gold* beads and biotin (*rose*), specifically binding avidin and related proteins (*blue*), which results in conductance changes [5]. (**b**) CPMV exposing streptavidin (SA, *pink*) and numerous fluorescent Cy5 dyes (*blue*) as signal enhancers in microarray detection technology. DNA-probe 1 (but not 2) is targeted by a biotinylated DNA (*red*). Its biotin molecules (*violet circles*) are captured by SA, which in 1:1 conjugation to Cy5 (*bottom*) does not yield a detectable signal (*right*), while it does on the viral 'Cy5 multiplyer' (*top/right*). [53] (**c**, **d**) *In situ* "bottom-up" assembly of TMV-like nucleoprotein tubes as multivalent templates on solid supports. (**c**) Bidirectional growth from CP oligo- and multimers on immobilized RNA, starting at the RNA's origin of assembly (OAs) structure; (**d**) spatially selectively equipped, patterned Si substrate with nanotubes on meandered areas fashioned with RNA (*via* DNA anchors coupled to aldehyde groups). AFM topography image [54] (reproduced from [5, 53, 54] with permission)

systems: Proof-of-concept experiments *e.g.* with CPMV [10, 48] confirmed that dual-functionalized viral shells with at least one binding moiety to a target, and multiple signal-generating units (fluorescent dyes, which might be replaced by quantum dots or further detectable molecules in other formats) were able to significantly enhance the sensitivity of analyte detection. To achieve this effect, the virus-derived functional additives were applied as secondary reactants, attaching to a hapten located on the primary probe (*e.g.* by means of streptavidin presented on a highly fluorescent virus, which thereby targeted biotin labels of nucleic acid hybridization probes or antibodies; refer to Fig. 22.9).

22.8.3 Drug Delivery and Diagnostics by Help of Viral Nanocontainers

The above-described possibilities arising from the incorporation of two or even more functionalities into individual viral capsids have drawn special attention to their putative use in diagnostic and therapeutic medical applications. Most appealing fields of utilization are the targeted delivery of indicator and effector substances to specific cells or tissues inside living organisms. While specific biogenic target structures may be found and bound by molecules presented on the outer viral capsid, the load of the viral cavity might serve as an imaging agent or, alternatively, as a drug. Plant and many bacterial and animal viruses are regarded ideal for such purposes due to their capsid dimensions, their proteinaceous multivalent surface (as described before), and their expected biocompatibility: They do not have any known pathogenic potential for humans and are usually tolerated without any notable adverse effects, probably due to their regular uptake with all types of food, ranging from vegetables and fruit to meat as well as seafood.

However, viral nanoparticles including plant-derived ones were shown to elicit immune responses and to penetrate many cell types and organs. Therefore, prior to any medical application, the risk of strong immune reactions and toxic effects due to high doses or specific modifications of viral capsids needs to be minimized. This is why numerous lines of research and development currently focus on strategies shielding the viral protein surfaces by covalent coatings with inert polymers, mainly polyethylene glycol, which not only reduces immunogenicity but obviously also increases the elimination rates (clearance) of viral capsids introduced into the body.

Before use, suitable preparations of viral nanocontainers intended to act as delivery vehicles for imaging agents or drugs (or genetic information in the case of gene therapeutical approaches) have to be loaded. Viral cavities may be filled with the respective cargo by diffusion (Sect. 22.6.1), supported by accordingly charged groups exposed inside, or by *in vitro* assembly of the viral capsid in the presence of the substance of choice (encapsulation). Some virus shells undergo conformational alterations upon changes of pH or ionic environment, which can lead to a convenient opening and closing of pores thus facilitating loading and retention of certain molecules. Once taken up into target tissues of treated organisms, however, cargo release into the surrounding cellular or intercellular medium may be difficult to achieve in a controlled manner, which is why novel externally inducible nanocontainer opening or destruction mechanisms are matter of intense current investigation (see Fig. 22.10 for a proposed strategy to achieve drug release by alternating electromagnetic fields).

Finally, two completely different applications of viral templates deserve to be touched in this section, although without being subjects of comprehensive description. One application undergoes powerful novel developments since viral capsids have been introduced into medical research as engineerable nanoparticulate effectors: the vaccination of animals and humans. Here, especially plants have shown an interesting potential to produce immunogenic, but non-infectious epitopes of animal pathogens or

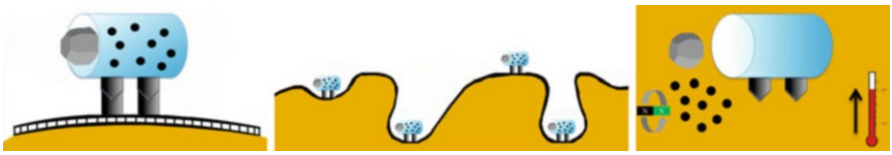


Fig. 22.10 Suggestion for a combination therapy against cancer, based on plant viruses. A plant virus is filled with a drug, which is thus shielded from blood and other cells. A modification with antibodies allows selective targeting of a tumor cell. The virus can enter the cell, and iron oxide nanoparticles (*large dots*) on or in the virus can be used for local hyperthermia, in combination with drug release (*small dots*). Hyperthermia requires an alternating magnetic field, applied externally. This and similar strategies are poised for breakthroughs in nanomedicine (image inspired by T. Pellegrino)

even tumor cells, either on the surface of chimeric plant virus particles, or as complete capsid structures in the case of heterologously assembled animal-viral particles. The resulting vaccines are devoid of any 'biorisk' of cell culture-derived contaminants. 'Edible' plant-derived mucosal vaccines against common diseases are under intense investigation, as well as plant viral particles equipped with peptide epitopes provoking immunogenic responses *e.g.* counteracting breast cancer [56]. The use of (engineered) VLPs and virions for vaccination has been described in detail in Chap. 21.

The last route of research presented here closes the circle back to the employment of viral templates in materials sciences: VLPs exposing peptides typical of extracellular matrix are used as novel functional coatings for tissue culture supports, thereby affecting adsorption and differentiation of animal and human cells growing on such substrates [57]. Those and related studies make use of the precise spacing of specific cell attachment signals if fused to viral CP subunits: By mimicking their spatial organization typical of certain organs, the viral hybrid particles are expected to help controlling the fate of interacting cells *in vitro*, with the aim of producing engineered tissue substitutes for a subsequent implantation into injured organs.

22.9 Other Systems

Mammalian viruses and their many uses for immunization and as gene vector could with full right be addressed as (rather advanced) nanotechnological instruments. This is however not usually done, due to their importance in medicine, and due to safety considerations, but probably also since most are more sensitive to environmental conditions than plant viruses (there are exceptions, though), and due to the small amounts available, compared with many plant viruses and phages, which can be obtained in grams. In any case, research and also applications in medicine are based on the infectivity for humans, but in nanotechnology this is always a disadvantage. However, enveloped viruses are used for arrays by exploiting their lipid layers.

A very good alternative are VLPs derived from animal viruses that can be prepared by removal of the RNA or DNA, or by assembly of pure CPs. Such nucleic acid-free virus shells are frequently employed for immunizations (see Chap. 21). For

the more complex animal virus particles (but not for simple viruses such as parvoviruses, nodaviruses and picornaviruses), assembly and surface chemistry may be technically more challenging than for simple plant viruses and phages. However, some specific properties of animal viruses and VLPs, such as their in-built capacity to recognize and penetrate specific cell types, including specific tumor cells, have made some of these virus particles (including engineered ones) relevant choices for the development of targeted gene therapy (since more than two decades), of drug delivery, and of other applications in biomedically-oriented nanotechnology. Details on the use of animal virus particles for biomedical and animal health applications are beyond the scope of this chapter, and the reader is referred to Chap. 21 regarding vaccine applications.

Archaea viruses attack archaea rather than bacteria, but many have a structure that is very similar to those of some phages. However, comparatively few are known, and their use in nanotechnology such as in combinatorial synthesis (phage display) is not established. The very high temperature stability of some archaeal viruses (especially hyper-thermophilic ones) may provide a stimulus for a broader use, though, and is interesting especially for the materials research side of nanotechnology.

Viroids are surprisingly simple infective agents, consisting only of an uncoated, tightly folded RNA molecule. Their discovery by Diener provided further stimulus in plant virology. In terms of physical properties and nanotechnology, they should be seen as a type of nucleic acid. While catalytic ribozyme sequences are very well known and used, viroids as a whole have not yet found applications in typical nanoscale science. With another view on nucleic acids, viral genomes feature many regulatory elements that have found use in heterologous protein production systems.

22.10 Perspectives and Conclusions

The challenges for virus nanotechnology are analogous to those for nanotechnology in general (Sect. 22.1). However, viruses imply the strict definition of nanoscale (< 100 nm), comparable to the smallest mass-produced structures. A major question is whether a (conventional) "top-down" or a bio-inspired "bottom-up" scheme is best suited. The answer has to be given for each structure separately; however, the smallest structures, of the size of viruses, might indeed require "bottom-up" and self-assembly strategies (Sect. 22.4). The use of viruses, especially for nonbiological nanotechnology (Sect. 22.1), requires "new surfaces", chemically or genetically engineered, and tailored for size and function. The background knowledge developed for such chemical modifications (Sects. 22.2 and 22.3) can also be used for templated structures (Sect. 22.6), and it is key for most applications (Sects. 22.7 and 22.8). The latter range from very small demonstration devices up to true production schemes, where also questions of biosafety and waste management would have to be addressed. Although much work has been started, the whole process of virus modification to achieve the desired functionality and/or properties is far from being

solved, and especially on the molecular scale (<5 nm) trial-and-error approaches are still prevalent over completely rational engineering. Key to overcome this problem are new analysis and manipulation tools (Sect. 22.5). Viruses, due to their high geometrical and chemical definition and functional properties, will then have the chance to become more generally involved in future nanotechnological breakthroughs.

Acknowledgements We are grateful for funding for the projects "Functionality-on-a-Stick" and "Serially ordered virus scaffolds" (Kompetenznetz "Funktionelle Nanostrukturen", Baden-Württemberg Stiftung), "MAGNIFYCO" (European Union, FP7-NMP4-SL-2009-228622), "Nanofluidica en Biotubos Moleculares" (Basque Country, PI2010-7), to the DFG, namely PAK410 and Priority Programmes SPP-1165 (Nanowires and nanotubes: from controlled synthesis to functions) and SPP-1569 (Generation of multifunctional inorganic materials by molecular bionics), and "NANOFLUID" (MICINN Spain, MAT2010-16184).

References and Further Reading

1. Lee S-Y, Lim J-S, Harris MT (2012) Synthesis and application of virus-based hybrid nanomaterials. Biotechnol Bioeng 109:16–30
2. Seeman NC (2010) Nanomaterials based on DNA. Annu Rev Biochem 79:65–87
3. Strable E, Finn MG (2009) Chemical modification of viruses and virus-like particles in viruses and nanotechnology. In: Manchester M, Steinmetz NF (eds) Current topics in microbiology and immunology, vol 327. Springer, Berlin/Heidelberg, pp 1–21
4. Douglas T, Young MJ (1998) Host–guest encapsulation of materials by assembled virus protein cages. Nature 393:152–155
5. Blum AS, Soto CM, Sapsford KE, Wilson CD, Moore MH, Ratna BR (2011) Molecular electronics based nanosensors on a viral scaffold. Biosens Bioelectron 26:2852–2857
6. Zhao Q, Chen W, Chen Y, Zhang L, Zhang J, Zhang Z (2011) Self-assembled virus-like particles from rotavirus structural protein VP6 for targeted drug delivery. Bioconjug Chem 22:346–352
7. Sarikaya M, Tamerler C, Schwartz DT, Baneyx F (2004) Materials assembly and formation using engineered polypeptides. Annu Rev Mater Res 34:373–408
8. Bruton J, Bouwer GT, Ward VK (2005) Metal nanoshell assembly on a virus bioscaffold. Nano Lett 5:1187–1191
9. Steinmetz NF, Evans DJ (2007) Utilisation of plant viruses in bionanotechnology. Org Biomol Chem 5:2891–2902
10. Steinmetz NF, Manchester M (2011) Viral nanoparticles – tools for materials science and biomedicine. Pan Stanford Publishing, Singapore
11. Miller RA, Presley AD, Francis MB (2007) Self-assembling light-harvesting systems. J Am Chem Soc 129:3104–3109
12. Endo M, Wang H, Fujitsuka M, Majima T (2006) Pyrene-stacked nanostructures. Chem Eur J 12:3735–3740
13. Pokorski JK, Steinmetz NF (2011) The art of engineering viral nanoparticles. Mol Pharmacol 8:29–43
14. Kadri A, Maiss E, Amsharov N, Bittner AM, Barlic S, Kern K, Jeske H, Wege C (2011) Engineered tobacco mosaic virus mutants with distinct physical characteristics in planta and enhanced metallization properties. Virus Res 157:35–46
15. Elliot S (1998) The physics and chemistry of solids. Wiley, Chichester
16. Jones RAL (2008) Soft machines: nanotechnology and life. Oxford University Press, Oxford

17. Israelachvili JN (2011) Intermolecular and surface forces, 3rd edn. Academic Press, Waltham
18. Insung SC, Bowden N, Whitesides GM (1999) Macroscopic, hierarchical, two-dimensional self-assembly. Angew Chem Int Ed 38:3078–3081
19. Lin Y, Su Z, Xiao G, Balizan E, Kaur G, Niu Z, Wang Q (2011) Self-assembly of virus particles on flat surfaces *via* controlled evaporation. Langmuir 27:1398–1402
20. Rong J, Lee LA, Li K, Harp B, Mello CM, Niu Z, Wang Q (2008) Oriented cell growth on self-assembled bacteriophage M13 thin films. Chem Commun 41:5185–5187
21. Lee S-W, Mao C, Flynn CE, Belcher AM (2002) Ordering of quantum dots using genetically engineered viruses. Science 296:892–895
22. Loo L, Guenther RH, Basnayake VR, Lommel SA, Franzen S (2006) Controlled encapsidation of gold nanoparticles by a viral protein shell. J Am Chem Soc 128:4502–4503
23. Kuznetsov YG, McPherson A (2011) Atomic force microscopy in imaging of viruses and virus-infected cells. Microbiol Mol Biol Rev 75:268–285
24. Young M, Willits D, Uchida M, Douglas T (2008) Plant viruses as biotemplates for materials and their use in nanotechnology. Annu Rev Phytopathol 46:361–384
25. Cuellar JL, Donath E (2012) Force microscopy – a tool to elucidate the relationship between nanomechanics and function in viruses. In: Frewin CL (ed) Atomic force microscopy investigations into biology – from cell to protein. Intech, Rijeka/Manhattan, pp 253–278
26. Guckenberger R, Heim M, Cevc G, Knapp HF, Wiegrabe W, Hillebrand A (1994) Scanning tunneling microscopy of insulators and biological specimens based on lateral conductivity of ultrathin water films. Science 266:1538–1540
27. Piner RD, Zhu J, Xu F, Hong S, Mirkin CA (1999) "Dip-pen" nanolithography. Science 283:661–663
28. Ashkin A, Dziedzic JM (1987) Optical trapping and manipulation of viruses and bacteria. Science 235:1517–1520
29. Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C (2001) The bacteriophage ϕ29 portal motor can package DNA against a large internal force. Nature 413:748–752
30. Alonso JM, Ondarçuhu T, Bittner AM (2013) Integration of plant viruses in electron beam lithography nanostructures. Nanotechnology, 24: 105305
31. Ermolina I, Milner J, Morgan H (2006) Dielectrophoretic investigation of plant virus particles: cowpea mosaic virus and tobacco mosaic virus. Electrophoresis 27:3939–3948
32. Green N, Morgan H, Milner J (1997) Manipulation and trapping of sub-micron bioparticles using dielectrophoresis. J Biochem Biophys Methods 35:89–102
33. Fraden S, Maret G, Caspar DLD (1993) Angular correlations and the isotropic-nematic phase transition in suspensions of tobacco mosaic virus. Phys Rev E 48:2816–2837
34. Hirai M, Koizumi M, Han R, Hayakawa T, Sano Y (2003) Right-/left-circular orientation of biological macromolecules under magnetic field gradient. J Appl Crystallogr 36:520–524
35. Tsukamoto R, Muraoka M, Seki M, Tabata H, Yamashita I (2007) Synthesis of CoPt and FePt3 nanowires using the central channel of tobacco mosaic virus as a biotemplate. Chem Mater 19:2389–2391
36. Kausche GA, Ruska H (1939) Die Sichtbarmachung der Adsorption von Metallkolloiden an Eiweißkörpern. Kolloid Z 89:21–26
37. Górzny MŁ, Walton AS, Evans SD (2010) Synthesis of high-surface-area platinum nanotubes using a viral template. Adv Funct Mater 20:1295–1300
38. Kobayashi M, Seki M, Tabata H, Watanabe Y, Yamashita I (2010) Fabrication of aligned magnetic nanoparticles using tobamoviruses. Nano Lett 10:773–776
39. Knez M, Bittner AM, Boes F, Wege C, Jeske H, Maiß E, Kern K (2003) Biotemplate synthesis of 3 nm nickel and cobalt nanowires. Nano Lett 3:1079–1082
40. Knez M, Sumser M, Bittner AM, Wege C, Jeske H, Martin TP, Kern K (2004) Spatially selective nucleation of metal clusters on the tobacco mosaic virus. Adv Funct Mater 14:116–124
41. Royston E, Ghosh A, Kofinas P, Harris MT, Culver JN (2008) Self-assembly of virus-structured high surface area nanomaterials and their application as battery electrodes. Langmuir 24:906–912

42. Nam KT, Kim DW, Yoo PJ, Chiang CY, Meethong N, Hammond PT, Chiang YM, Belcher AM (2006) Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. Science 312:885–888
43. Miura A, Hikono T, Matsumura T, Yano H, Hatayama T, Uraoka Y, Fuyuki T, Yoshii S, Yamashita I (2006) Floating nanodot gate memory devices based on biomineralized inorganic nanodot array as a storage node. Jpn J Appl Phys 45:L1–L3
44. Górzny MŁ, Walton AS, Wnek M, Stockley PG, Evans SD (2008) Four-probe electrical characterization of Pt-coated TMV-based nanostructures. Nanotechnology 19:165704
45. Tseng RJ, Tsai C, Ma L, Ouyang J, Ozkan CS, Yang Y (2006) Digital memory device based on tobacco mosaic virus conjugated with nanoparticles. Nat Nanotechnol 1:72–77
46. Wu Z, Mueller A, Degenhard S, Ruff E, Geiger F, Bittner AM, Wege C, Krill C III (2010) Enhancing the magnetoviscosity of ferrofluids by the addition of biological nanotubes. ACS Nano 4:4531–4538
47. Lee BY, Zhang J, Zueger C, Chung W-J, Yoo SY, Wang E, Meyer J, Ramesh R, Lee S-W (2012) Virus-based piezoelectric energy generation. Nat Nanotechnol 7(6):351–356
48. Soto CM, Ratna BR (2010) Virus hybrids as nanomaterials for biotechnology. Curr Opin Biotechnol 21:426–438
49. Werner S, Marillonnet S, Hause G, Klimyuk V, Gleba Y (2006) Immunoabsorbent nanoparticles based on a tobamovirus displaying protein A. Proc Natl Acad Sci USA 103:17678–17683
50. Smolenska L, Roberts IM, Learmonth D, Porter AJ, Harris WJ, Wilson TMA, Santa Cruz S (1998) Production of a functional single chain antibody attached to the surface of a plant virus. FEBS Lett 441:379–382
51. Montague NP, Thuenemann EC, Saxena P, Saunders K, Lenzi P, Lomonossoff GP (2011) Recent advances of cowpea mosaic virus-based particle technology. Hum Vaccines 7:383–390
52. Souza GR, Christianson DR, Staquicini FI, Ozawa MG, Snyder EY, Sidman RL, Miller JH, Arap W, Pasqualini R (2006) Networks of gold nanoparticles and bacteriophage as biological sensors and cell-targeting agents. Proc Natl Acad Sci USA 103:1215–1220
53. Soto CM, Blum AS, Vora GJ, Lebedev N, Meador CE, Won AP, Chatterji A, Johnson JE, Ratna BR (2006) Fluorescent signal amplification of carbocyanine dyes using engineered viral nanoparticles. J Am Chem Soc 128:5184–5189
54. Mueller A, Eber FJ, Azucena C, Petershans A, Bittner AM, Gliemann H, Jeske H, Wege C (2011) Inducible site-selective bottom-up assembly of virus-derived nanotube arrays on RNA-equipped wafers. ACS Nano 5:4512–4520
55. Mao C, Liu A, Cao B (2009) Virus-based chemical and biological sensing. Angew Chem Int Ed 48:6790–6810
56. Frolova OY, Petrunia IV, Komarova TV, Kosorukov VS, Sheval EV, Gleba YY, Dorokhov YL (2010) Trastuzumab-binding peptide display by tobacco mosaic virus. Virol 407:7–13
57. Lee LA, Nguyen QL, Wu L, Horvath G, Nelson RS, Wang Q (2012) Mutant plant viruses with cell binding motifs provide differential adhesion strengths and morphologies. Biomacromolecules 13:422–431

Further Reading

Bäeuerlein E, Behrens P, Epple M (2007) Handbook of biomineralization. Wiley, Weinheim
Jones RAL (2008) Soft machines: nanotechnology and life. Oxford University Press, Oxford
Steinmetz NF, Manchester M (2011) Viral nanoparticles – tools for materials science and biomedicine. Pan Standford Publishing, Singapore

Also especially recommended for further reading are references [1–3, 7, 9, 24, 48, 51] *listed above*.

Index

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8,

T

Index of virus species cited in the text

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8,

Index of virus families cited in the text

M.G. Mateu (ed.), *Structure and Physics of Viruses: An Integrated Textbook*, Subcellular Biochemistry 68, DOI 10.1007/978-94-007-6552-8,