Subcellular Biochemistry 96

J. Robin Harris
Jon Marles-Wright *Editors*

Macromolecular Protein Complexes III: Structure and Function

Springer

Subcellular Biochemistry

Volume 96

The book series SUBCELLULAR BIOCHEMISTRY is a renowned and well recognized forum for disseminating advances of emerging topics in Cell Biology and related subjects. All volumes are edited by established scientists and the individual chapters are written by experts on the relevant topic. The individual chapters of each volume are fully citable and indexed in Medline/Pubmed to ensure maximum visibility of the work.

More information about this series at http://www.springer.com/series/6515

J. Robin Harris · Jon Marles-Wright
Editors

Macromolecular Protein Complexes III: Structure and Function

Springer

Editors
J. Robin Harris
Institute of Molecular Physiology
University of Mainz
Mainz, Rheinland-Pfalz, Germany

Jon Marles-Wright
School of Natural and Environmental Sciences
Newcastle University
Newcastle upon Tyne, UK

ISSN 0306-0225 ISSN 2542-8810 (electronic)
Subcellular Biochemistry
ISBN 978-3-030-58970-7 ISBN 978-3-030-58971-4 (eBook)
https://doi.org/10.1007/978-3-030-58971-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume of the Subcellular Biochemistry series devoted to Macromolecular Protein Complexes continues the theme of Volumes 83 and 94, by presenting a diverse selection of interesting chapter topics within which emphasis is placed primarily upon soluble protein complexes, together with some membrane- and virus-related protein topics. These two latter areas were previously covered in Volumes 87 and 88 of the series, but will be integrated, along with fibrous proteins, in the present and future volumes of the series in an almost encyclopaedic manner. The recently published Volume 94, entitled *Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins* also complements the content of the present volume.

The compilation and editing of advanced multi-author scientific books of this nature is a difficult task in the present state of demanding academic life and intense scientific activity, where professional advancement is very much linked to publication of primary research outputs within prestigious scientific journals rather than books. Nevertheless, the book format continues to provide an opportunity for authors to present their own research achievements and express their opinions in a freer, and more expansive manner than is possible within review articles. With Open Access options now available for book chapters, together with electronic download, this gives equally broad distribution and readership possibilities to the scientific journals.

Chapter topics in this book were selected from the recently published research work of authors who have interesting data available for full presentation and integration with the research data of other groups. The structural and functional aspects presented in these chapters are enhanced with views on the developing biotechnological, biomedical, and pharmaceutical potential of the complexes discussed. The 18 chapter topics included in the book, listed immediately following in the Contents list, will not be expanded upon here, apart from stating that grouping of topics has not been performed here, owing to the diverse nature of the chapter content. Some do link together, but in the main all are free-standing. Of significance is the lengthy and important presentation on the complex structural and functional dynamics of the 26S proteasome. Unfortunately, from our initial list of agreed

chapters five good topics were lost during the extended production period, by the authors failing to submit their material! This is perhaps a reflection of the extreme pressures of work on those concerned and the unusual circumstances we find ourselves in with the global COVID-19 pandemic.

We hope that this volume will provide broadly useful information for structural biologists, biochemists and biomedical scientists, ranging from the advanced undergraduate level, through to postgraduate and career scientists.

Mainz, Germany — J. Robin Harris
Newcastle upon Tyne, UK — Jon Marles-Wright

Contents

Chapter 1
Structure, Dynamics and Function of the 26S Proteasome

Youdong Mao

Abstract The 26S proteasome is the most complex ATP-dependent protease machinery, of ~2.5 MDa mass, ubiquitously found in all eukaryotes. It selectively degrades ubiquitin-conjugated proteins and plays fundamentally indispensable roles in regulating almost all major aspects of cellular activities. To serve as the sole terminal "processor" for myriad ubiquitylation pathways, the proteasome evolved exceptional adaptability in dynamically organizing a large network of proteins, including ubiquitin receptors, shuttle factors, deubiquitinases, AAA-ATPase unfoldases, and ubiquitin ligases, to enable substrate selectivity and processing efficiency and to achieve regulation precision of a vast diversity of substrates. The inner working of the 26S proteasome is among the most sophisticated, enigmatic mechanisms of enzyme machinery in eukaryotic cells. Recent breakthroughs in three-dimensional atomic-level visualization of the 26S proteasome dynamics during polyubiquitylated substrate degradation elucidated an extensively detailed picture of its functional mechanisms, owing to progressive methodological advances associated with cryogenic electron microscopy (cryo-EM). Multiple sites of ubiquitin binding in the proteasome revealed a canonical mode of ubiquitin-dependent substrate engagement. The proteasome conformation in the act of substrate deubiquitylation provided insights into how the deubiquitylating activity of RPN11 is enhanced in the holoenzyme and is coupled to substrate translocation. Intriguingly, three principal modes of coordinated ATP hydrolysis in the heterohexameric AAA-ATPase motor were discovered to regulate intermediate functional steps of the proteasome, including ubiquitin-substrate engagement, deubiquitylation, initiation of substrate translocation and processive substrate degradation. The atomic dissection of the innermost working of the 26S proteasome opens up a new era in our understanding of the ubiquitin-proteasome system and has far-reaching implications in health and disease.

Y. Mao (✉)
Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Harvard Medical School, Boston 02215, Massachusetts, USA
e-mail: youdong_mao@dfci.harvard.edu

School of Physics, Center for Quantitative Biology, Peking University, Beijing 100871, China

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_1

Keywords Proteasome · Ubiquitylation · Ubiquitin-proteasome system · Deubiquitylation · AAA-ATPase motor · Proteolysis · Mechanochemistry · Homeostasis · Cryogenic electron microscopy · Cryo-EM · Conformational dynamics

Introduction

Human cells express more than 20,000 genes, which direct synthesis of a comparable number of proteins by the ribosome. Intracellular constituents dynamically preserved as proteins are continuously synthesized and degraded in cells. Newly synthesized proteins might misfold into dysfunctional structures that are cytotoxic and must be removed in a timely manner. Many initially well-folded proteins may need to be partially degraded or cleaved to activate specific function. Functional proteins may ultimately become unnecessary once they fulfil their missions. All these proteolytic needs can be met by the master recycling machinery called the 26S proteasome (Ciechanover and Kwon 2015; Mayer 2000; Meyer-Schwesinger 2019; Hnia et al. 2019; Coux et al. 1996; Finley and Prado 2019; Bard et al. 2018; Finley et al. 2016; Collins and Goldberg 2017; Voges et al. 1999; Tanaka 2009). The 2.5-MDa 26S proteasome counteracts the ribosome, by controlling the fate of synthesized proteins, and tightly regulates intracellular protein contents (Livneh et al. 2016; Collins and Goldberg 2017). To date, the 26S proteasome remains the largest, most enigmatic degradation machinery known and is ubiquitously found in all eukaryotic kingdoms (Darwin 2009; Muller and Weber-Ban 2019).

To be degraded by the proteasome holoenzyme, the targeted proteins must first be covalently modified by ubiquitin moieties via a cascade of three types of enzyme, namely, the E1 ubiquitin-activating enzyme, the E2 ubiquitin-conjugating enzyme, and the E3 ubiquitin ligase (Ciehanover et al. 1978; Ciechanover et al. 1980a, b; Hershko et al. 1979, 1980). Ubiquitin is a highly conserved protein with 76 amino acids ubiquitously found in eukaryotic cells (Wilkinson et al. 1980; Goldstein et al. 1975). Regulated modification of protein substrates with ubiquitin initiates their recognition and breakdown by the proteasome in an ATP-dependent fashion (Etlinger and Goldberg 1977; Arrigo et al. 1988; Wilkinson et al. 1980; Hershko et al. 1980). Mammalian cells contain only very few E1 enzymes, several E2 enzymes and at least several hundreds of different E3 ligases (Buetow and Huang 2016; Zheng and Shabek 2017; Hua and Vierstra 2011). The hierarchical cascade of ubiquitylation enzymes allows orthogonal regulation of hundreds of substrates in parallel. A large network of proteins involved in ubiquitylation and ubiquitin-mediated degradation by the 26S proteasome constitutes the ubiquitin-proteasome system (UPS) (Mayer 2000; Ciechanover 2005; Finley and Prado 2019; Livneh et al. 2016). Through the UPS, the proteasome elegantly regulates all major aspects of cellular processes, such as the cell cycle, gene expression, signal transduction, immune response, apoptosis and carcinogenesis (Ciechanover 2005). To serve as the sole terminal "processor" for myriad ubiquitylation pathways, the proteasome evolved exceptional

adaptability in dynamically organizing ubiquitin receptors, shuttle factors, deubiquitinases, AAA-ATPase unfoldases, ubiquitin ligases and a large network of proteins to enable substrate selectivity, processing efficiency and precision of intracellular regulation (Finley and Prado 2019). The proteasome holoenzyme is one of the most dynamic enzymes known. The inner working of the proteasome is among the most sophisticated, enigmatic mechanisms of enzymatic machinery.

Indeed, owing to extremely complex dynamics in the 26S proteasome, the elucidation of its atomic structures took about four decades following its first discovery in the late 1970s (Etlinger and Goldberg 1977) despite continuous efforts by the worldwide research community and several generations of scientists (Bard et al. 2018). Building upon recent cryogenic electron microscopy (cryo-EM) studies on the human proteasome dynamics (Frank 2006; Chen et al. 2016a; Lu et al. 2017b; Zhu et al. 2018), an eventual breakthrough in visualizing substrate-processing dynamics of the human 26S proteasome at the atomic level (Dong et al. 2019) revealed the inner working of this sophisticated holoenzyme for the first time, clarified many long-term outstanding problems in the mechanisms by which a polyubiquitylated substrate is recognized, deubiquitylated, unfolded and degraded by the proteasome. These discoveries were partly echoed or complemented by separate structural studies of the substrate-bound yeast proteasome (de la Pena et al. 2018) and of other related systems in lower organisms (Puchades et al. 2017; Yu et al. 2018; Cooney et al. 2019; Majumder et al. 2019; Puchades et al. 2019; Twomey et al. 2019; Ripstein et al. 2020; Ding et al. 2019). These studies together establish the "central dogma" of the proteasome, thus revolutionizing our mechanistic understanding of the UPS. Comparisons among these studies further provide insights into conservation and variation of the operating principles used by proteolytic machineries in general. In this chapter, I offer a thorough account for our current understanding on the structure, dynamics and function of the 26S proteasome. Throughout, all descriptions are presented in the context of the human 26S proteasome in default, for brevity, unless otherwise explicitly stated.

Overview of the Ubiquitin-Proteasome System

Ubiquitylation

Ubiquitin and its polymerized form serve as the degradation signals for substrate recognition by the 26S proteasome (Komander and Rape 2012). Ubiquitin is an essential member in the protein family that shares a globular β-grasp fold of around 70 amino acids with a surface-exposed C-terminal glycine residue. Canonical ubiquitylation of a protein substrate is achieved via an isopeptide bond connecting the C-terminal glycine of ubiquitin with the ε-amino group of any lysine side chain exposed on the solvent-accessible surface of the substrate (Ciehanover et al. 1978,

1980a, b; Hershko et al. 1979, 1980). In rare non-canonical cases, covalent attachment of ubiquitin to cysteine, threonine, serine or the N-terminal amino group in the substrate have also been observed (Kravtsova-Ivantsiv and Ciechanover 2012).

Ubiquitylation involves sequential, cascaded actions of the ubiquitin-activating enzyme (E1), the ubiquitin-conjugating enzyme (E2), and the ubiquitin ligase (E3), and the ubiquitin chain-elongation enzyme (E4) families (Fig. 1.1) (Koegl et al. 1999). The E1 enzyme catalyzes ATP-dependent activation of ubiquitin and isopeptide bond formation. The activated ubiquitin is then transferred to an E2 enzyme. The activated intermediate of E2-ubiquitin functions as the ubiquitin donor to the E3 ligase enzyme, which binds both the substrate and E2-ubiquitin intermediate via different structural motifs and recruits the ubiquitylation target as the substrate for the proteasome (Baek et al. 2020). Four families of E3 ligases have been discovered and classified into HECT, U-box, RING and RING-between-RING (RBR), based on their molecular architecture and enzymatic mechanisms. To enable selectivity and diversity in substrate processing, eukaryotic cells have evolved at least hundreds of distinct E3 ligases that feature a broad spectrum of substrate-interacting regions resembling a limited number of structural scaffolds (Buetow and Huang 2016; Zheng and Shabek 2017; Hua and Vierstra 2011). For example, the multi-subunit cullin-RING ligases (CRLs) in the RING family of E3 enzymes employ one of several cullin isoforms to assemble the complex (Baek et al. 2020). The remarkable diversity of the E3 enzymes allows the UPS to operate in various intracellular contexts for tightly regulating specific cellular activities via proteasomal degradation (Samant et al. 2018). In some cases, the first set of ubiquitin molecules attached to a protein substrate is further ubiquitylated so that the substrate becomes modified with longer polyubiquitin chains. The additional ubiquitin moieties can be attached by the same E3 or by a different ubiquitin ligase E4 (Koegl et al. 1999) that can only add more ubiquitin to substrate-conjugated monoubiquitin or polyubiquitin chains such as UBE3C/Hul5 and Ufd2 (Crosas et al. 2006; Hanzelmann et al. 2010).

Many substrates of the proteasome are ubiquitylated on more than one lysine residue. The ubiquitylation site on a substrate itself is not tightly bound to the E3 so that structural flexibility of the ubiquitylation sites would allow several different lysine residues to interact with the ubiquitin bound to the E2 in the E2-E3-substrate intermediate complex (Baek et al. 2020). Through the conjugation cascade, the substrates could be modified with only one ubiquitin, with several ubiquitin molecules at distinct sites, or with polyubiquitin chains that are concatenated covalently via any surface-exposed lysine residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63) or the N-terminal α-amino group of methionine (Met1) in ubiquitin. These three types of ubiquitin conjugation are referred to as monoubiquitylation, multiubiquitylation and polyubiquitylation, respectively (Kirisako et al. 2006; Xu et al. 2009; Yau et al. 2017). With the growth of the polyubiquitin chains, branched or tree-like heterotypic chain topologies have been observed (Yau et al. 2017). Such a complexity in ubiquitin polymerization generates highly diverse, heterogeneous degradation signals and allows ubiquitylation to regulate myriad cellular functions, some of which are not directly linked to proteolysis (Oh et al. 2018). The complexity of the ubiquitin signals is reversely "decoded" through the proteasome-dependent

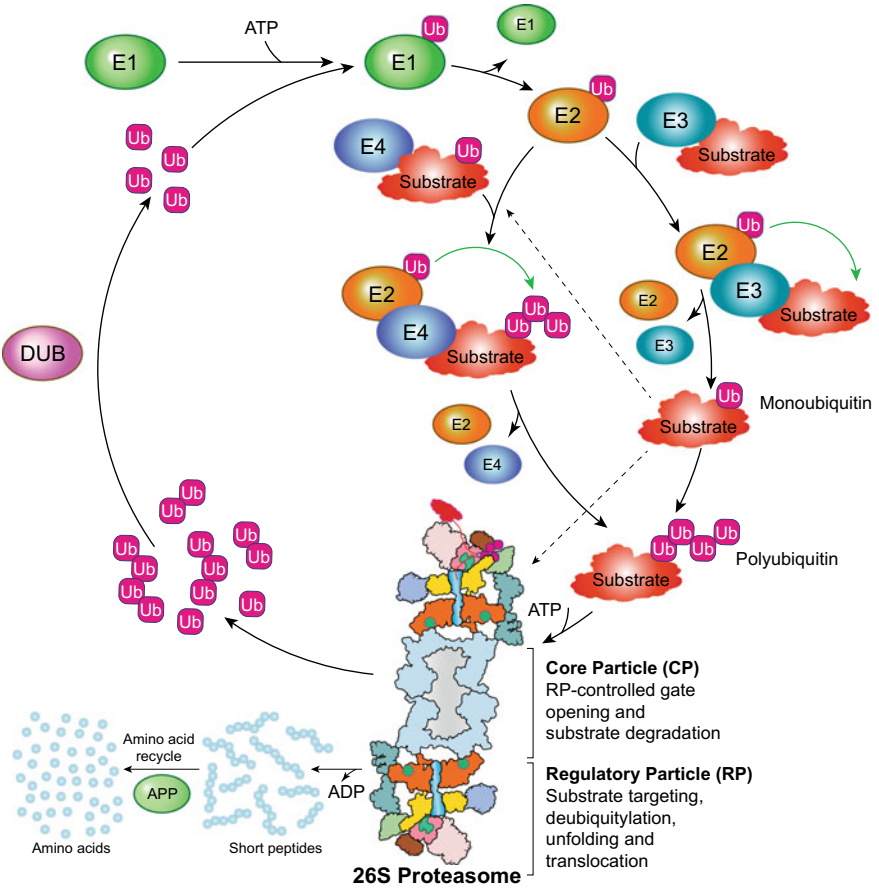


Fig. 1.1 Overview of the ubiquitin-proteasome system. Ubiquitylation is accomplished by the covalent conjugation of ubiquitin to lysine residues exposed on the surface of targeted substrates. Ubiquitin is first activated by the E1 in the presence of ATP. The ubiquitin is then transferred from E1 to E2. A ubiquitin ligase (E3) recruits the ubiquitin-bound E2 enzyme and a substrate to transfer the ubiquitin from E2 to the substrate. A specific type of ubiquitin ligase (E4) elongates the ubiquitin chains after monoubiquitin or polyubiquitin. The formation of Lys48-linked polyubiquitin chains, the main signal for proteasomal degradation, is mediated by successive cycles of ubiquitin conjugation. Monoubiquitylation could be sufficient for proteasome targeting in certain scenario. The monoubiquitin and polyubiquitin are removed from the substrate by the deubiquitinases (DUBs) associated with the proteasome. After being translocating into the proteasome, substrates are cleaved into short peptides, which are further broken down to amino acids by aminopeptidases (APPs). Released free polyubiquitin molecules are further recycled by cytosolic DUBs for another round of ubiquitylation

recruitment of several ubiquitin receptors (Husnjak and Dikic 2012; Lu et al. 2015; Oh et al. 2018), and many deubiquitylating enzymes (DUBs) capable of selectively processing substrates with various types of ubiquitin linkages (de Poot et al. 2017; Clague et al. 2019; Komander et al. 2009).

26S Proteasome Holoenzyme

As the sole proteolytic "central processor" in the UPS, the 26S proteasome holoenzyme assembles through noncovalent association of two subcomplexes, core particle (CP) and regulatory particle (RP). The RP caps either or both sides of the CP cylinder (Fig. 1.2). The CP, also known as the 20S proteasome, ubiquitously exists in prokaryotes and eukaryotes (Coux et al. 1996). The RP subcomplex, also known as PA700 or 19S, is a vital, indispensable proteasome activator found in all eukaryotic cells. The RP recruits, deubiquitylates, unfolds and translocates protein substrates into the CP chamber for degradation in an ATP-dependent fashion. In archaeal cells, one of the equivalent ATP-dependent activators is the proteasome-activating nucleotidase (PAN) complex, an ortholog of the eukaryotic proteasomal AAA-ATPase (Benaroudj et al. 2003; Smith et al. 2005). Several ATP-independent activators, typically PA28

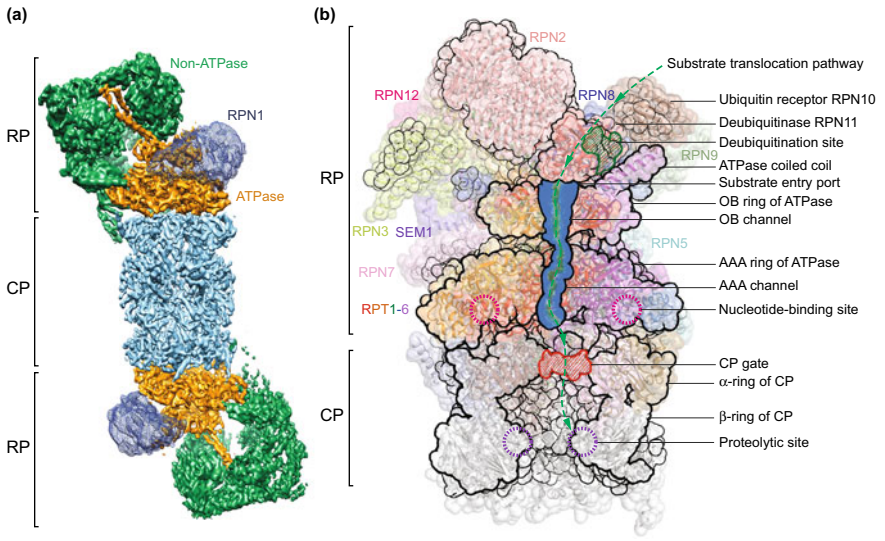


Fig. 1.2 Overall architecture of the 26S proteasome. **a** The cryo-EM reconstruction of the doubly capped 26S proteasome. The CP is shown as cyan. The RP is marked by the dashed box. **b** Illustrative anatomy of the RP-CP subcomplex structure, with black silhouettes representing the central cross-section along the ATPase channel colored opaque blue. Dashed green curve illustrates the substrate-translocation pathway across the channel to the proteolytic sites in the CP chamber. Only upper half of the CP is shown

(11S/REG) and PA200 (Blm10), can also bind the CP for proteasome activation (Stadtmueller and Hill 2011).

Core Particle

Proteasomal peptidase activities are housed in the CP. The CP is composed of 14 α-type and 14 β-type protein subunits (Lowe et al. 1995; Groll et al. 1997; Arrigo et al. 1988; Dong et al. 2019). In eukaryotic cells, there are seven distinct α-type or β-type subunits. Each type of subunit assembles into a heteroheptameric ring. Four such rings stack into a cylinder-shaped barrel in an $\alpha_{1-7}\beta_{1-7}\beta_{1-7}\alpha_{1-7}$ arrangement with an approximately two-fold symmetry (Lowe et al. 1995; Groll et al. 1997). In prokaryotic cells, there is only one gene for either the α-type or β-type subunit that assembles into a homoheptameric ring. Similar to the proteasomal CP, several other proteolytic proteins also assemble into a barrel-shaped proteolytic chamber, such as HslV (Bochtler et al. 1997) and ClpP (Wang et al. 1997). In eukaryotes, each β-ring houses three proteolytically active threonine residues in β1, β2 and β5 and forms the catalytic chamber. The active sites of the peptidase are sequestered inside the catalytic chamber. Heptameric α-rings, positioned on both sides of the catalytic chamber, control substrate entry into this space. Opening of an axial channel within the α-ring is allosterically controlled by the interactions of the CP with the 19S RP (PA700) or other activators such as PA200 and PA28 (Whitby et al. 2000; Zhu et al. 2018; Dong et al. 2019).

In lymphoid tissues or other interferon-γ (IFN-γ) stimulated cells, three constitutively expressed β-subunits are substituted with the subtype subunits β1i, β2i and β5i to assemble a variation of the CP named the immunoproteasome (Ferrington and Gregerson 2012). Different subtypes of β subunits may be mixed and combined in a single proteasome assembly to form an intermediate-type CP that exhibits different enzymatic properties (Dahlmann et al. 2000). Other CP subtypes such as thymus-specific (Murata et al. 2007) and testis-specific CPs (Uechi et al. 2014) have been observed to contain subunits β5t and α4s, which replace β5 and α4, respectively. These cell-type specific CPs were also termed the thymoproteasome (β5t) and spermatoproteasome (α4s), respectively (Murata et al. 2007, 2018). Subtype variations of the CP fine-tune its catalytic activity to meet tissue-specific functional needs.

Regulatory Particle

The RP consists of at least 18 protein subunits and regulates substrate recognition, recruitment, and unfolding in an ATP-dependent manner. The RP is structurally divided into two subcomplexes named the lid and the base that may assemble separately (Glickman et al. 1998). The lid subcomplex comprises 9 Regulatory Particle Non-ATPase (RPN) subunits, i.e., RPN3 (PSMD3/S3), RPN5 (PSMD12), RPN6 (PSMD11/S9), RPN7 (PSMD6/S10), RPN8 (PSMD7/S12), RPN9 (PSMD13/S11), RPN11 (PSMD14/Poh1/Pad1), RPN12 (PSMD8/S14), and

RPN15 (PSMD9/Dss1/Sem1) (Table 1.1). Six subunits (RPN3, RPN5, RPN6, RPN7, RPN9 and RPN12) exhibits similar architecture featuring an N-terminal Proteasome-CSN-Initiation factor 3 (PCI) domain (Hofmann and Bucher 1998). Two subunits (RPN8 and RPN11) feature an MPR1-PAD1 N-terminal (MPN) domain and form an MPN dimer (Tran et al. 2003). All of the eight subunits are decorated with a C-terminal α-helical domain separated via a structured linker from their core PCI/MPN domains. The lid subcomplex is organized into a horseshoe-like architecture through an elaborate bundle that is composed of the C-terminal α-helical domain of each lid subunit except RPN15 (Estrin et al. 2013).

The base subcomplex includes RPN1 (PSMD2/S2), RPN2 (PSMD1/S1), and six paralogous, distinct Regulatory Particle ATPase (RPT) subunits—RPT1 (PSMC2/S7), RPT2 (PSMC1/S), RPT3 (PSMC4/S6), RPT4 (PSMC6/S10), RPT5 (PSMC3/S6a) and RPT6 (PSMC5/S8)—from the classic ATPases Associated with diverse cellular Activities (AAA) family (Table 1.1). The six RPT subunits of the base share a general domain organization, consisting of an N-terminal coiled-coil (CC) domain, an oligonucleotide- and oligosaccharide-binding (OB) domain, and a C-terminal AAA domain. They form a heterohexameric ATPase ring that acts as a mechanical motor. The well-folded domains of a substrate are unfolded by the mechanical force produced by the ATPase motor ring and translocated via the central pore of the ATPase ring by harvesting the chemical energy of ATP hydrolysis catalyzed by the ATPases. Thus, the ring-like RPT hexamer is also referred to as an unfoldase or translocase in the literature (Bard et al. 2018).

The first step of substrate processing by the proteasome is recognition of a ubiquitylated substrate, which is mediated by the ubiquitin receptors within the base, including RPN1 (PSMC2/S7) (Shi et al. 2016), RPN10 (PSMD4/S5a) (Deveraux et al. 1994; van Nocker et al. 1996), and RPN13 (ADRM1) (Husnjak et al. 2008; Schreiner et al. 2008). RPN13 can reversibly associate with RPN2 and is sub-stoichiometric in the endogenously purified human proteasome. In addition to the ubiquitin receptors intrinsically residing in the proteasome, ubiquitylated substrates can also be recruited and delivered to the proteasome by extrinsic ubiquitin receptors consisting of ubiquitin-like (UBL) and ubiquitin-associated (UBA) domains, including RAD23, DSK2, and DDI1 (Elsasser and Finley 2005; Elsasser et al. 2004; Zhang et al. 2009a). These UBL-UBA proteins interact with the intrinsic ubiquitin-binding sites in the proteasome via their UBL domain, functioning as shuttle factors that diversify the routes of decoding the ubiquitin signals for substrate selection and recognition.

During substrate engagement, an unstructured initiation region of the substrate is recognized by the pore loops of the RPT subunits (Prakash et al. 2004; Yu and Matouschek 2017; Bard et al. 2019; Dong et al. 2019). To allow substrate translocation into the CP, conjugated ubiquitin chains are cleaved from substrates by either the DUB RPN11 (Verma et al. 2002; Yao and Cohen 2002; Worden et al. 2017; Dong et al. 2019) or other auxiliary DUBs like USP14 (ubiquitin-specific protease 14) (Lee et al. 2016) and UCH-L5 (ubiquitin carboxyl-terminal hydrolase isozyme L5) (Vander Linden et al. 2015). RPN11 is a zinc-dependent DUB in the JAB1/MPN/MOV34 (JAMM) family protein (Ambroggio et al. 2004) and is intrinsically located above

the OB ring, guarding the entry of substrates into the ATPase ring and CP. It catalyzes *en bloc* substrate deubiquitylation when a substrate is threaded into the ATPase ring.

The axial entrance of the CP proteolytic chamber, named the CP gate, is closed in the resting state (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016). Opening of the CP gate is allosterically triggered by insertion of the C-termini of all RPT subunits except RPT4 into the outward pockets of the α-ring (α-pockets) located between adjacent α-subunits (Chen et al. 2016a; Zhu et al. 2018; Dong et al. 2019; Eisele et al. 2018; de la Pena et al. 2018; Smith et al. 2007; Rabl et al. 2008). The PA200 and PA28 complexes are two alternative CP-activating regulators that can replace the 19S RP in an assembled holoenzyme (Table 1.1). They trigger CP gate opening in an ATP-independent manner due to lack of ATPase domain. PA200 is a monomeric protein of ~250 kDa and conserved from the yeast to human (Schmidt et al. 2005; Blickwedehl et al. 2008). The PA28 family is expressed in higher eukaryotes and some unicellular eukaryotes such as trypanosomes. In mammalian cells, three isoforms (PA28α, PA28β, PA28γ) form two kinds of activators with distinct properties of expression, localization and activation. PA28α and PA28β are 28-kDa proteins that assemble into heteroheptameric ring structures in vertebrates (Ma et al. 1992; Dubiel et al. 1992) (Table 1.1).

History of Proteasome Structure Determination

X-Ray Crystallography of Proteasome Components

It has been half century since the first image of the human CP was recorded by negative-stain electron microscopy when its function was unknown (Fig. 1.3) (Harris 1968). The gradual elucidation of the proteasome structure and function was propelled and accompanied by innovations in a broad set of methods and tools in biochemistry and structural biology. Because the proteasomal CP is highly stable overall, it became the first key component to be crystallized. In the 1990s, both atomic structures of archaeal and yeast 20S proteasome were solved by X-ray crystallography (Lowe et al. 1995; Groll et al. 1997). The crystal structure of the human CP was solved much later (Harshbarger et al. 2015; Schrader et al. 2016). Crystal structures of the isolated CP or immunoproteasome CP are available for several species (Groll et al. 1997; Unno et al. 2002; Harshbarger et al. 2015; Huber et al. 2012). These structural works defined three conserved proteolytic sites in the β1, β2 and β5 subunits, laying an important foundation for understanding the activation of the CP and the mechanism of proteolysis. However, all these high-resolution CP structures represent a basal resting state with a closed gate in their α-rings. The first yeast CP with an open gate in the α-ring was determined by X-ray crystallography in the presence of activation by the PA26 (11S) regulator (Whitby et al. 2000), whereas the activated human CP in an open-gate state was only determined much later by cryo-EM (Chen et al. 2016a; Zhu et al. 2018; Dong et al. 2019).

Table 1.1 Nomenclature, structure and function of proteasome subunits and proteasome-associated proteins

Subcomplex	Systematic nomenclature	Human genome name	Other mammalian name	Yeast name	Structure	Function
CP (20S) α-ring	α1	PSMA6	iota	Scl1, YC7	α/β/β/α sandwich, NLS	RP docking
	α2	PSMA2	C3	Pre8, Y7	α/β/β/α sandwich, NLS	CP gating, RP docking
	α3	PSMA4	C9	Pre9, Y13	α/β/β/α sandwich, NLS	CP gating, RP docking
	α4	PSMA7	C6	Pre6	α/β/β/α sandwich	CP gating, RP docking
	α5	PSMA5	zeta	Pup2, DOA5	α/β/β/α sandwich	RP docking
	α6	PSMA1	C2	Pre5	α/β/β/α sandwich	RP docking
	α7	PSMA3	C8	Pre10, YC1	α/β/β/α sandwich	
CP (20S) β-ring	β1	PSMB6	Y, delta	Pre3	α/β/β/α sandwich, Ntn	Caspase-like activity
	β2	PSMB7	Z	Pup1	α/β/β/α sandwich, Ntn	Trypsin-like activity
	β3	PSMB3	C10	Pup3	α/β/β/α sandwich	
	β4	PSMB2	C7	Pre1	α/β/β/α sandwich	
	β5	PSMB5	X, MB1, epsilon	Pre2, DOA3	α/β/β/α sandwich, Ntn	Chymotrypsin-like activity
	β6	PSMB1	C5	Pre7	α/β/β/α sandwich	
	β7	PSMB4	N3, beta	Pre4	α/β/β/α sandwich	

(continued)

Table 1.1 (continued)

Subcomplex	Systematic nomenclature	Human genome name	Other mammalian name	Yeast name	Structure	Function
Immunoproteasome	β1i	PSMB9	LMP2, RING12	–	α/β/β/α sandwich, Ntn	Chymotrypsin-like activity
	β2i	PSMB10	MECL-1, LMP10	–	α/β/β/α sandwich, Ntn	Trypsin-like activity
	β5i	PSMB8	LMP7, RING10	–	α/β/β/α sandwich, Ntn	Chymotrypsin-like activity
Thymoproteasome	β5t	PSMB11		–	α/β/β/α sandwich, Ntn	Chymotrypsin-like activity
RP (PA700, 19S) Base	RPT1	PSMC2	S7, Mss1, p48	Rpt1, CIM5, YTA3	AAA	Unfoldase
	RPT2	PSMC1	S4, p56	Rpt2, YHS4, YTA5, mts2	AAA	Unfoldase
	RPT3	PSMC4	S6, Tbp7, p47	Rpt3, YNT1, YTA2	AAA	Unfoldase
	RPT4	PSMC6	S10b, p42	Rpt4, CRL13, PCS1, SUG2	AAA	Unfoldase
	RPT5	PSMC3	S6', Tbp1	Rpt5, YTA1	AAA	Unfoldase
	RPT6	PSMC5	S8, P45, Trip1	Rpt6, CIM3, SUG1, let1	AAA	Unfoldase
	RPN1	PSMD2	S2, P97	Rpn1, HRD2, NAS1, mts4	Toroidal LRR, PC	Ub/UBL receptor
	RPN2	PSMD1	S1, P112	Rpn2, SEN3	Toroidal LRR, PC, NLS	RPN13 docking
	RPN10	PSMD4	S5a, p54	Rpn10, MCB1, SUN1	VWA, UIM	Ub/UBL receptor

(continued)

Table 1.1 (continued)

Subcomplex	Systematic nomenclature	Human genome name	Other mammalian name	Yeast name	Structure	Function
	RPN13	ADRM1		Rpn13, DAQ1	PRU	Ub/UBL receptor
RP (PA700, 19S) Lid	RPN3	PSMD3	S3, p58	Rpn3, SUN2	PCI, PAM	Scaffold
	RPN5	PSMD12	p55	Rpn5, NAS5	PCI	AAA regulation
	RPN6	PSMD11	S9, p44.5	Rpn6, NAS4	PCI, PAM	AAA regulation
	RPN7	PSMD6	S10a, p44	Rpn7	PCI	AAA regulation
	RPN8	PSMD7	S12, p40, MOV34	Rpn8	MPN	Assisting deubiquitylation
	RPN9	PSMD13	S11, p40.5	Rpn9, NAS7	PCI	Scaffold
	RPN11	PSMD14	S13, Poh1	Rpn11, MPR1	MPN, DUB	Deubiquitinase
	RPN12	PSMD8	S14, p31	Rpn12, NIN1	PCI	Scaffold
	RPN15	PSMD9	S15, p27L, DSS1	Sem1, HOD1	Helical	Scaffold
PA28 (11S REG)	PA28α	PSME1	PA28α, REGα	–		CP activator
	PA28β	PSME2	PA28β, REGβ	–		CP activator
	PA28γ	PSME3	PA28γ, REGγ, Ki	–		CP activator
PA200	PA200	PSME4	PA200, TEMO	Blm10	HEAT, ARM	CP activator
CP-associated proteins	PAC1	PSMG1		Pba1, Poc1		α-ring chaperone
	PAC2	PSMG2		Pba2, Poc2, ADD66		α-ring chaperone
	PAC3	PSMG3		Pba3, Poc3, Dmp2		α-ring chaperone
	PAC4	PSMG4		Pba4, Poc4, Dmp1		α-ring chaperone
	UMP1	POMP	hUmp1, Proteassemblin	Ump1		β-ring chaperone

(continued)

Table 1.1 (continued)

Subcomplex	Systematic nomenclature	Human genome name	Other mammalian name	Yeast name	Structure	Function
RP-associated proteins	RAD23A/B	RAD23A/B	hHR23A/B	Rad23	UBL, UBA	Extrinsic Ub receptor
	UBQLN1/2/3/4	UBQLN1/2/3/4	hPLIC-1/2/3/4, ubiquilin-1/2/3/4	Dsk2	UBL, UBA	Extrinsic Ub receptor
	DDI1	DDI1		Ddi1	UBL, UBA	Extrinsic Ub receptor
	ZFAND5	ZFAND5	ZNF216		Zinc-finger	Extrinsic Ub receptor
	USP14	USP14		Ubp6	UBL, DUB	Deubiquitjnase
	UCH37	UCHL5	UCH37	Uch37, Uch2	UCH, DUB	Deubiquitinase
	UBE3C	UBE3C	KIAA10	Hul5	HECT	E3 ligase
	Parkin	PARK2			UBL, RBR	E3 ligase
	ECM29	KIAA0368		Ecm29	HEAT repeats	Regulation of proteasome assembly
	PI31	PSMF1			HbYX motif, Proline-rich	Inhibition of proteasome assembly
	NAS6	PSMD10	p28, gankyrin	Nas6	Ankyrin repeats	RP assembly chaperone
	NAS2	PSMB9	p27, BRIDGE, Rac2	Nas2	PDZ	RP assembly chaperone
	HSM3	PSMD5	S5b, Rac3, p50.5	Hsm3	HEAT repeats	RP assembly chaperone
	RPN14	PAAF1	FLJ11848	Rpn14, YGL004C	WD-40	RP assembly chaperone

AAA: ATPase associated with diverse cellular activities. ARM: Armadillo repeats. DUB: Deubiquitylating enzyme. HEAT: Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1. HECT: a domain homologous to the E6AP carboxyl terminus. HbYX: hydrophobic-tyrosine-X. MPN: Mpr1/Pad1/N-terminal. NLS: Nuclear localization signal. Ntn: N-terminal nucleophile hydrolase. PAC: Proteasome assembling chaperone. PAM: PCI-associated module. PCI: proteasome/COP9/eIF3. PDZ: PSD-95/DLG/ZO-1. PRU: Pleckstrin-like receptor for ubiquitin. RBR: Ring-between-ring. UBA: Ubiquitin associated. UBL: Ubiquitin-like. UIM: Ubiquitin-interacting motif. VWA: von Willebrand factor type A

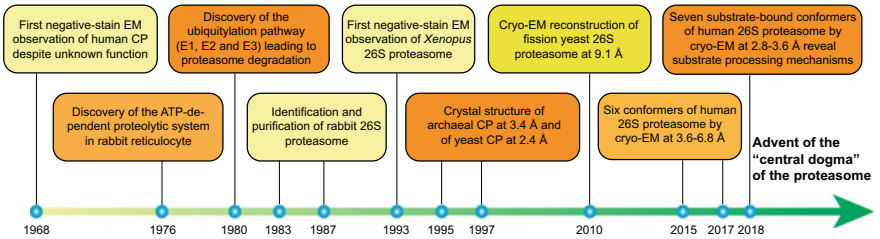


Fig. 1.3 Timeline and milestones in the elucidation of structure and function of the 26S proteasome. In 1968, the first image of the human proteasome CP was captured by negative-stain EM when there was no knowledge about its function (Harris 1968). In the 1970s, concerted discoveries revealed the functional role of the proteasome by studying ATP- and ubiquitin-dependent degradation system in eukaryotic cells (Etlinger and Goldberg 1977; Ciehanover et al. 1978; Hershko et al. 1979; Ciechanover et al. 1980a, b; Hershko et al. 1980; Wilkinson et al. 1980). In the 1980s, the approaches of purification of mammalian 26S proteasome were developed, paving the way to its structural analysis (Tanaka et al. 1983; Hough et al. 1986, 1987). In 1993, the first image of the mammalian 26S proteasome was captured by negative-stain EM (Peters et al. 1993). Soon after, X-ray crystallography revealed the proteasome CP structures in both Archaea and yeast (Lowe et al. 1995; Groll et al. 1997). In 2010, the first subnanometer resolution structure of a yeast 26S proteasome was reconstructed by cryo-EM (Bohn et al. 2010). During 2015–2017, the first high-resolution structures of the substrate-free human proteasome were determined by cryo-EM, which revealed six coexisting conformations (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016; Zhu et al. 2018), whereas cryo-EM structures of their yeast counterparts were also determined at the similar resolution range in the following years (Eisele et al. 2018; Wehmer et al. 2017; Ding et al. 2017, 2019). In 2018, the first atomic structures of the substrate-engaged human 26S proteasome in seven functional states were determined to higher resolution by cryo-EM (Dong et al. 2019), whereas the yeast counterparts in four substrate-engaged states were analyzed at slightly lower resolutions (de la Pena et al. 2018)

In stark contrast to conformational stability of the CP, the structure of the RP is highly dynamic and samples an extensively expanded conformational landscape (Lu et al. 2017b; Unverdorben et al. 2014; Chen et al. 2016a; de la Pena et al. 2018; Zhu et al. 2018; Dong et al. 2019). Thus, the RP structure in the 26S proteasome has completely evaded X-ray crystallography. However, many but not all of the RP subunits were separately solved in certain domains or regions by X-ray crystallography or nuclear magnetic resonance (NMR) (He et al. 2012; Pathare et al. 2012; Boehringer et al. 2012; Riedinger et al. 2010; Worden et al. 2014; Pathare et al. 2014; Zhang et al. 2009b, c; Schreiner et al. 2008; Shi et al. 2016). Despite its simplicity, the archaeal PAN complex, an ATPase homohexamer, is thought to represent an evolutionary precursor to the eukaryotic proteasomal ATPases. The OB and AAA domains of the PAN ATPase have been separately crystallized and solved in a fragmented form (Zhang et al. 2009b, c).

Early Electron Microscopy Studies of Proteasome at Low Resolution

The 19S RP in the 26S proteasome is highly dynamic and prevents the 26S proteasome from being crystallized. Thus, electron microscopy reconstruction became the only option to obtain the molecular shape of the 26S proteasome in the 1990s (Peters et al. 1993; Frank 2006). The first cryo-EM reconstruction of the 26S proteasome from *Drosophila* at 1–2 nm resolution was documented in 2009 (Nickell et al. 2009). Following the improvement of single-particle reconstruction method, the cryo-EM structure of substrate-free *Schizosaccharomyces pombe* 26S proteasome at a subnanometer resolution (9.1 Å) was reported (Bohn et al. 2010). This level of resolution was insufficient to allow direct identification of subunit organization. Thus, other methods in biochemistry and cross-linking mass spectrometry were used in an integrative fashion to define the subunit architecture of the RP (Beck et al. 2012; Lander et al. 2012; Lasker et al. 2012; da Fonseca et al. 2012). Using the crystal structures of individual RP subunits, the first pseudo-atomic model for the yeast 26S proteasome was built using ~ 8 Å maps (Beck et al. 2012). In the following years, additional cryo-EM reconstructions of the yeast proteasome at subnanometer resolutions were analyzed under various substrate-free conditions and revealed three distinct conformational states (designated s1, s2 and s3), which were hypothetically interpreted as the substrate-accepting, commitment and translocating states, respectively (Matyskiela et al. 2013; Unverdorben et al. 2014; Sledz et al. 2013). Meanwhile, a 9-Å map of the substrate-bound yeast proteasome was reconstructed but the density of substrate inside the proteasome was averaged out (Matyskiela et al. 2013). Nevertheless, the overall conformation of the substrate-bound yeast proteasome resembles that of state s3, although they are not completely identical. The limited resolution, however, precludes understanding of the structural mechanisms for ATP-dependent degradation by the 26S proteasome (Forster et al. 2013). Inter-subcomplex regulation, coordinated ATP hydrolysis in the AAA-ATPases, and substrate-proteasome interactions remained particularly elusive in these studies.

High-Resolution Cryo-EM Studies of Substrate-Free Proteasome

My colleagues and I started working on the human proteasome structure in 2014. Our first near-atomic resolution reconstruction of proteasome was determined in the late 2015 (Chen et al. 2016a). However, it showed closed gates on both sides of the CP, in contradiction to early conclusions derived from the lower resolution cryo-EM maps (Beck et al. 2012; Lander et al. 2012; Lasker et al. 2012; da Fonseca et al. 2012). To clarify this issue, additional analysis was conducted to test the hypothesis on a possible coexisting conformational state with an open CP gate. A focused classification strategy was devised by subtracting the CP density from the original

raw single-particle cryo-EM images, following which exhaustive 3D classification eventually sorted out four coexisting conformations (designated states S_A, S_B, S_C and S_D) of the human proteasome in presence of ATP/Mg^{2+}, among which only one state S_D was open in the gate of CP (Chen et al. 2016a). This allowed us to observe that insertion of five RPT C-terminal tails into the CP albeit at a moderate resolution (6–8 Å) for the first time (Chen et al. 2016a). The cryo-EM structure of the human proteasome in state S_A but no other states were also reported at comparable resolutions by two other groups (Huang et al. 2016; Schweitzer et al. 2016). In the following year, a high-resolution structure of the yeast proteasome in state s1 (analogous to human S_A) was reported, along with an open CP state s4 (analogous to human S_D) at ~8-Å resolution discovered when the yeast proteasome was bound to ATP analogs instead of ATP (Wehmer et al. 2017).

Following the first reconstruction of human proteasome at the atomic level, several groups set out to determine the atomic structures of the substrate-bound proteasome holoenzyme. However, the exceptional dynamics prevent the complex from being reconstructed at high resolution by using the same cryo-EM procedure practiced in the determination of the substrate-free proteasome in the resting state (S_A). It was soon realized that unprecedented challenges in cryo-EM and biochemical methodology must be conquered before one can obtain an atomic structure of the substrate-bound proteasome. First, how can one capture the complex before substrate degradation is completed? Second, how can one deconvolute the exceptional conformational heterogeneity in the sample not otherwise commonly encountered? Third, how can one push as many proteasome conformations as possible to the near-atomic resolution range (2.5–3.6 Å)? To confront these formidable challenges and to find a feasible path forward, we devised two pilot studies along with several parallel cryo-EM methodology developments (Xu et al. 2016; Zhu et al. 2017; Wu et al. 2017; Wang et al. 2019), with the anticipation of gaining new ideas for devising novel methods for solving the substrate-bound proteasome at the atomic level.

In one study, the free RP complex was analyzed by cryo-EM in states presumably prior to its assembly with the CP. In the free RP, the AAA-ATPase ring exhibited continuous motion of a broad conformational landscape, captured in seven conformational states at ~9 Å resolution in the AAA-ATPase but 4.6 Å for the rest of the free RP (Lu et al. 2017b). In another study, the 26S proteasome was analyzed in the presence of ATPγS that replaces ATP. It was aimed to improve the open-gate state to high resolution (Zhu et al. 2018). Surprisingly, ATPγS induced a dramatic shift in the conformational landscape of the proteasome from state S_A to S_D. To improve the resolution of state S_D, we had to further classify the dataset corresponding to state S_D into three distinct sub-states S_{D1}, S_{D2} and S_{D3}, which were then improved to 4.3–4.9 Å (Zhu et al. 2018). Averaging of the three sub-states instead gave rise to a 6-Å resolution with smeared features in the center of the ATPase ring. These studies taught us an important lesson that 3D classification of a highly heterogeneous dataset with significantly improved clustering accuracy is crucial for achieving higher resolution for those lowly populated conformational states. The observations of six coexisting states of the ATPγS-bound human proteasome was soon echoed by the

discovery of states s5 and s6 of the yeast proteasome (Eisele et al. 2018), confirming the conservation of the proteasome dynamics in the absence of a substrate.

Visualizing Atomic-Level Dynamics of Functional Proteasome

In the early summer of 2018, seven distinct states of the substrate-engaged proteasome at 2.8–3.6 Å resolution (Fig. 1.4), which are referred to as E_{A1}, E_{A2}, E_B, E_{C1}, E_{C2}, E_{D1} and E_{D2}, eventually emerged after a few years of marathon-like repeated protein purification, data collection and analysis (Dong et al. 2019), in sharp contrast to some of our other cryo-EM structures solved in a month (Zhang et al. 2015). These exciting results would not be made possible without several lines of methodology innovation and improvement. First, instead of completely replacing ATP with ATPγS or nucleotide analogs like several studies (Sledz et al. 2013; Wehmer et al. 2017; Eisele et al. 2018; Zhu et al. 2018; Yu et al. 2018), a novel "nucleotide-substitution" strategy that dilutes ATP with ATPγS in a time-dependent manner was deliberately devised to capture the functional proteasome conformations in action that are as diverse as possible (Dong et al. 2019). The proteasome was first primed with a bona fide polyubiquitylated substrate $Sic1^{PY}$ and ATP/Mg^{2+} for 30 s, then exchanged in a buffer containing both ATP and ATPγS at a 1:1 stoichiometric ratio (Dong et al. 2019). The choice of 30 s delay in ATPγS dilution after substrate-proteasome mixing was based on the approximate half-life of the substrate degradation reaction, which is consistent with a recent kinetic measurement on proteasomal degradation (Bard et al. 2019). This strategy decelerated the hydrolysis activity of AAA-ATPases by partially replacing ATP with ATPγS in the proteasome. This is expected to maximally preserve the conformational landscape and native heterogeneity of the substrate-engaged proteasome being captured before the completion of substrate degradation reactions (Dong et al. 2019) (Fig. 1.4).

Second, to overcome the difficulties of cryo-EM data analysis entailed by an extreme degree of conformational heterogeneity in this case, we decided to collect a considerably larger cryo-EM dataset (44,664 raw micrographs) that is necessary

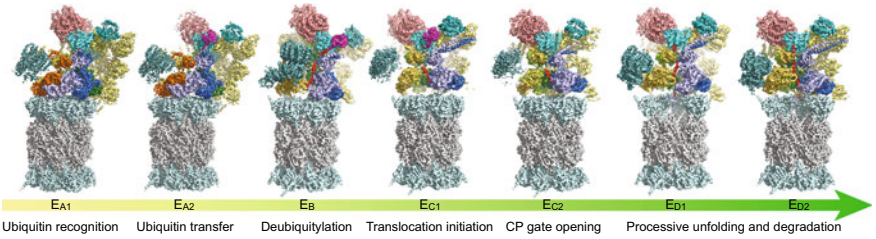


Fig. 1.4 Seven cryo-EM maps of the substrate-engaged *Homo sapiens* 26S proteasome at 2.8–3.6 Å resolution capture the key intermediate steps of substrate processing and provide important insights into the chemical cycle of proteasome-mediated degradation (Dong et al. 2019)

for improving resolutions for those lowly populated conformational states (Dong et al. 2019), which would otherwise be either invisible or of low resolution with a normally sized dataset. Note that a much larger dataset would drastically complicate the computational task of 3D classification rather than straightforwardly ensuring higher classification accuracy or higher resolution, because it could readily break the upper limit of high-performance computational resources, causing frequent failure of popular cryo-EM software. In fact, more data can effectively reduce resolution rather than increasing it if the data classification accuracy is not warranted (Wu et al. 2017).

Third, we extensively used the latest data processing tools developed in my laboratory, including deep-learning-based particle selection (Zhu et al. 2017), manifold-learning-based 2D clustering (Wu et al. 2017), and hardware-accelerated 3D classification to optimize the data processing. Without these new computational tools, it could have taken at least another year or longer to analyze the same dataset. Furthermore, building upon lessons learnt from previous studies, we further fine-tuned a hierarchically focused 3D classification strategy with stepwise zoomed-in masking and eventually sorted out seven conformations of the substrate-bound proteasome covering the nearly complete cycle of substrate processing from substrate recruitment and deubiquitylation to processive degradation (Dong et al. 2019).

Before fully committed to the "nucleotide substitution" strategy, there was a debate in my group regarding whether we should instead follow a more conventional, lower-risk strategy of stabilizing the substrate-bound proteasome by inhibiting the deubiquitylating activity of RPN11, which was previously used in a cryo-EM study of the substrate-bound yeast proteasome (Matyskiela et al. 2013). But concerns were that this treatment might be insufficient to lock the proteasome into a single conformation, could fail in capturing the proteasome in functional steps other than substrate translocation and also potentially create some artificial structural features that are not physiologically relevant. These concerns turned out to be mostly the case when four alternative cryo-EM structures of the substrate-bound yeast proteasome at 4.2–4.7 Å resolution were obtained by inactivating the yeast Rpn11 with the inhibitor orthophenanthroline (de la Pena et al. 2018). The inhibition of the yeast Rpn11 results in certain structural features in state 1D* that are potentially off-pathway. However, the AAA-ATPase structures in states 5D, 5T and 4D of the substrate-engaged yeast proteasome are remarkably comparable to their human counterparts in states E_{D1} and E_{D2} within the limit of their reconstruction quality and resolution (de la Pena et al. 2018; Dong et al. 2019). This suggests that the molecular mechanisms of substrate processing by the proteasome must be highly conserved throughout the eukaryotic kingdom.

With hindsight, the proteasome has served as an extraordinary model system or "gold-standard trial" for furthering methodology development in cryo-EM for its potential in solving atomic-level dynamics of complex enzymes (Dong et al. 2019). By retrospective comparison, it appears that the structural determination of functional substrate-engaged proteasome was perhaps much more difficult than those of the ribosome and spliceosome, the other two well-known dynamic protein complexes. The conventional cryo-EM methodology presumes a homogeneous core

structure or few dominating conformations that can afford high-accuracy image alignment leading to high-resolution reconstructions. However, the substrate-engaged proteasome samples a significantly expanded, highly frustrated energy landscape and severely violates the presumption of the conventional cryo-EM methodology. Misclassification of conformers was found to be the major cause for limited resolutions when the proteasome samples an expanded energy landscape (Zhu et al. 2018). However, improved curation in initial particle ensemble with deep learning, higher accuracy in data clustering using manifold learning, a fine-tuned, object-optimized 3D classification strategy and a significantly larger dataset all seem to contribute to the improvement of 3D classification accuracy beyond the realm of the conventional cryo-EM methodology. A more detailed account for the methodology advancement leading to the breakthrough of atomic structure determination of functional human proteasome will be provided elsewhere.

Structures and Dynamics of the Proteasome

Molecular Architecture of the Proteasome

Due to the highly dynamic nature of the proteasome, no single reconstructions of the 26S proteasome provided the best resolved local features for all components (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016; Zhu et al. 2018; Dong et al. 2019; de la Pena et al. 2018; Eisele et al. 2018; Ding et al. 2019). Overall, the CP and AAA-ATPase motor in the RP were best resolved in state E_A at 2.8-Å resolution; the non-ATPase subunits in the RP were best resolved in states E_{D1} and E_{D2} at 3.3 and 3.2 Å (Dong et al. 2019). The early high-resolution structures of the proteasome in state S_A all presented poorer local density quality in the lid and the non-ATPase subunits in the base, where the local resolution did not go beyond the range of 4–8 Å (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016; Zhu et al. 2018). This issue was addressed in the recent study at higher resolution in these subunits (Dong et al. 2019). Putting together, these atomic structures provide thus-far the nearly complete understanding of the 26S proteasome assembly. In this subsection, I intended to summarize the overall organization and architecture of the proteasome holoenzyme based on the highest resolution structures available. More details about conformational dynamics of the proteasome associated with distinct functional steps are discussed in the following subsections.

The Lid Subcomplex

The ~370-kDa lid subcomplex in the RP has an overall dimension of 170 × 150 × 140 $Å^3$. The architecture of the ~370-kDa lid subcomplex in the RP bears topological similarity to those of the COP9 signalosome (Lingaraju et al. 2014) and

the eukaryotic translation initiation factor 3 (eIF3) (des Georges et al. 2015). The PCI domain of six RPN subunits (RPN9, RPN5, RPN6, RPN7, RPN3, and RPN12 from the left to right side in Fig. 1.5a) features N-terminal helical repeats (HR) connected to a winged-helix (WH) subdomain. The lid subcomplex is stabilized by two organizational centers (Fig. 1.5): a half-sized open ring formed through side-by-side oligomerization of three-stranded β-sheets of the WH subdomains from the six PCI proteins (PCI ring) (Fig. 1.5a), and an elaborate α-helical bundle comprising the C-terminal regions from all of eight PCI/MPN subunits (Fig. 1.5b). The N-terminal helical repeats of the six PCI proteins radially protrude from the PCI ring of 18-stranded β-sheet, forming a horseshoe-like architecture.

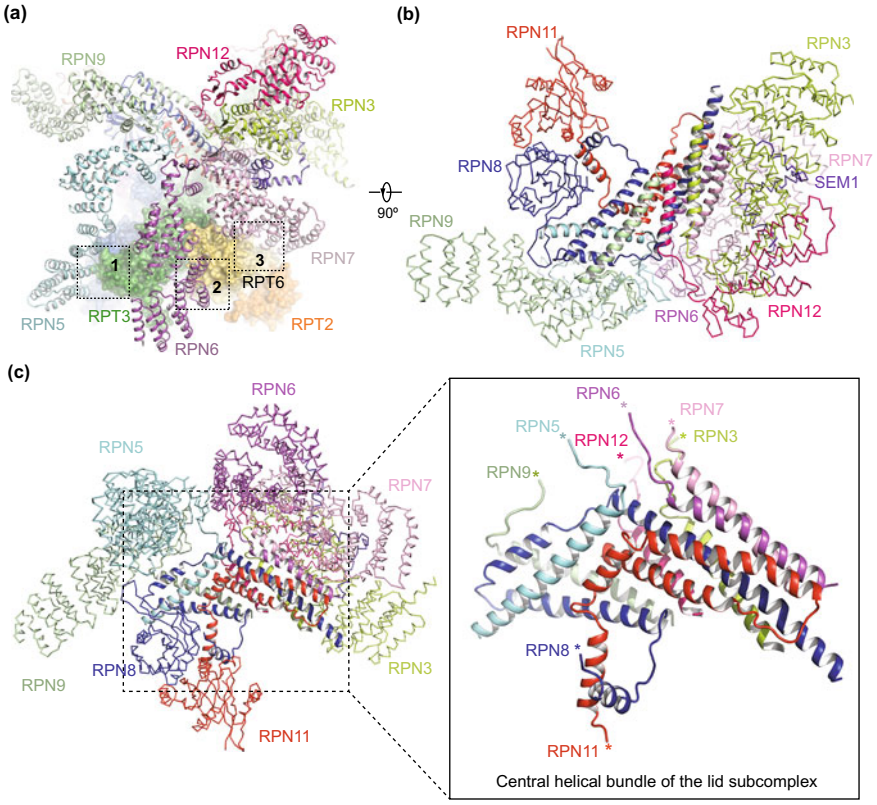


Fig. 1.5 Atomic structures of the lid subcomplex in the human 26S proteasome (PDB ID: 6MSB) (Dong et al. 2019). **a** The lid-base interface lateral to the ATPase ring is composed of three local inter-subunit interfaces of RPN5-RPT3, RPN6-RPT6 and RPN7-RPT6, which are highlighted by dashed boxes 1, 2 and 3, respectively. **b**, **c** The atomic model of the lid subcomplex viewed from two perspectives orthogonal to that in (**a**). The central helical bundle is highlighted in cartoon representation. The rest of the lid structure is shown in a ribbon representation. The right inset zooms into the central helical bundle. The asterisk symbols label the N-terminal ends of the lid subunits

The PCI and MPN proteins form largely two separate subassemblies that are united in the central helical bundle. The central helical bundle is connected to the PCI ring from one side and to the MPN dimer of RPN8-RPN11 from the other side via structured linkers at the C-termini of the WH subdomains. Outside the helical bundle, the MPN dimer does not directly contact the PCI ring. It is separated by the base subunit RPN2 from one end and the von Willebrand factor type A (VWA) domain of another non-lid subunit RPN10 from the other end. Only the MPN domain of isopeptidase RPN11 in the lid subcomplex harbors a catalytic active zinc ion.

The Base Subcomplex

The base subcomplex is organized around the ATPase ring (RPT1-RPT6). The coiled-coil (CC) dimers of RPT1-RPT2, RPT3-RPT6, and RPT4-RPT5 harbor RPN1, RPN2 and RPN10, respectively. While the interaction between the RPT3-RPT6 CC domain and RPN2 was apparent in the cryo-EM maps of the proteasome in the resting state (S_A), the other two pairs of interactions were insufficiently visualized due to the local conformational dynamics of RPN1 in the absence of a substrate (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016; Zhu et al. 2018). The interface between RPT1-RPT2 CC and RPN1 was best visualized in the substrate-bound proteasome in states E_{D1} and E_{D2} at higher resolution (Dong et al. 2019).

RPN10 was originally thought to be part of the base (Glickman et al. 1998), but later considered to be neither part of the base nor the lid (Bard et al. 2018), based on the structural observation that the RPN10 VWA domain interacts directly with the lid subunits RPN8 and RPN9 in the proteasome and does not contact any base subunits in the resting state of the proteasome (Chen et al. 2016a). Nevertheless, cryo-EM structures of substrate-interacting proteasome did show a direct interaction of RPN10 with the base at the RPT4-RPT5 CC in state E_{D2} (Dong et al. 2019). Furthermore, the RPT4-RPT5 CC dimer seems to harbor auxiliary ubiquitin-binding sites that assist ubiquitin recognition by RPN10 and peptide-proximal ubiquitin transfer to RPN11 for deubiquitylation (Dong et al. 2019; Chen et al. 2020; Lam et al. 2002). The ubiquitin-interacting motifs (UIM) of RPN10 are structurally flexible and missing in all available cryo-EM reconstructions (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016; Zhu et al. 2018; Dong et al. 2019). In the yeast proteasome, Rpn13 binds the C-terminal region of Rpn2 with a sub-stoichiometric ratio (Sakata et al. 2012).

The OB domains of six RPT subunits form a ring with an inner diameter of 10 Å (Fig. 1.6). This is the substrate entry port into the interior of the proteasome. Its narrow dimension indicates that the OB ring imposes the restriction against translocation of folded substrates. Each of the OB domains folds into a five-stranded β-barrel, whose axis lies orthogonal to that of the OB ring and is directed radially (Fig. 1.6h). Residues in loop L23 appear to impose the narrowest constriction in the central pore of the OB ring subcomplex. Despite being surrounded by highly dynamic structures, the OB ring mostly moves as a whole rigid body during ATP or substrate turnover.

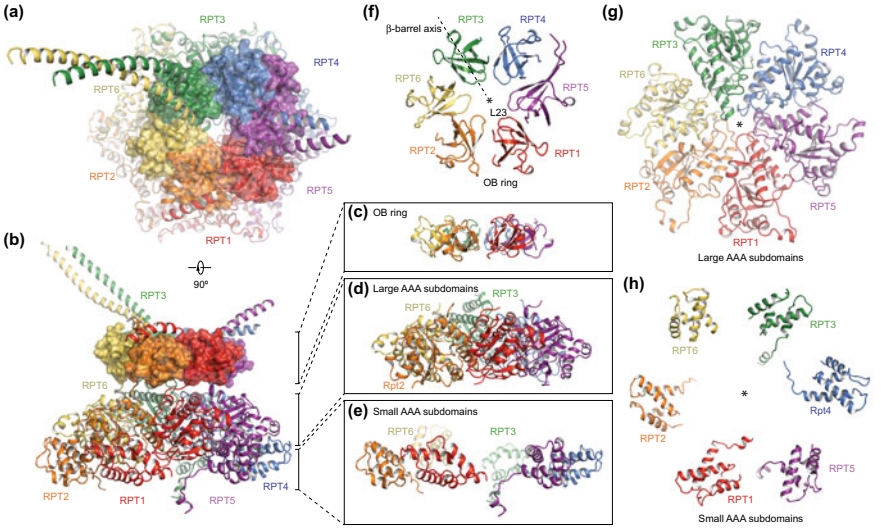


Fig. 1.6 The atomic model of the ATPase heterohexamer in the E_A/S_A state (PDB ID: 6MSB) (Dong et al. 2019). **a** The overall structure of the ATPase viewed from the perspective of the substrate entry port. The OB ring is in the foreground and is shown with transparent surface superimposed on a cartoon representation of the AAA domains. **b** Overall structure of the ATPase ring viewed from a perspective rotated 90° from that in panel (**c**). **c–e** The component structures of the OB ring (panel **c**), large AAA subdomains (panel **d**) and small AAA subdomains (panel **e**) viewed from the same perspective as that in panel (**b**). **f–h** The component structures of the OB ring (panel **f**), large AAA subdomains (panel **g**) and small AAA subdomains (panel **h**) viewed from the same perspective as that in panel (**a**)

The link between the CC and OB domains is marked by a conserved proline residue in the *cis* configuration in RPT2 (Pro106), RPT3 (Pro86) and RPT5 (Pro87) at homologous residue positions in the structure. Thus, a *cis*-Pro residue in every other subunit is essential for dimerization of the CC domains of three RPT heterodimers (RPT1-RPT2, RPT3-RPT6 and RPT4-RPT5). Alternating *cis*-Pro residues at this position have also been found in the archaeal PAN homohexamer—a remarkable instance of evolutionary conservation (Zhang et al. 2009b; Arcus 2002).

The AAA domains of RPT subunits have a common architecture, featuring an N-terminal α/β subdomain (large AAA subdomain) linked to a smaller C-terminal α-helical subdomain (small AAA subdomain) via a short loop. The concave pocket between the large and small AAA subdomain houses the nucleotide-binding site. In the resting state (S_A) of the proteasome, the AAA domains of RPT subunits form a spiral staircase resembling a lock washer, in which RPT3 is located at the top position and RPT2 at the bottom, with RPT6 bridging RPT2 and RPT3 at the "split seam". RPT6 possesses the most divergent fold in its AAA domain, consistent with its unique role in driving the initial substrate engagement (Dong et al. 2019).

The Lid-Base Interface

The interfaces between the lid and the base play crucial roles in coordinating substrate recognition, deubiquitylation and translocation (Greene et al. 2019). The lid-base interface is organized around the DUB RPN11, a metalloprotease fold, which is positioned at the substrate entrance of the OB ring of the ATPase heterohexamer. On one side of RPN8-RPN11, the MPN dimer is demarcated from RPN2, and the other side from the VWA domain of RPN10. There are three types of lid-base interfaces: (1) the central interface between RPN11 and the CC-OB domains of the ATPase ring; (2) the interface defined by the CC domains of the adjacent RPT3 and RPT6 subunits, which project distally from the central interface; and (3) the lateral interface between the AAA domains of ATPases and RPN5-RPN7. The interfaces between the lid and the ATPase ring bury a ~3900 $Å^2$ surface area in total. Specifically, the CC-OB domains of RPT3-RPT6 contribute a ~3100 $Å^2$ interfacial area with the lid. The RPT3-RPT6 CC domain is encircled by the helices in RPN2, RPN3, RPN8 and RPN11. This interfacial architecture appears to stabilize the lid-base association.

The lateral lid-base interface implicates distinct roles of interacting subunits in the conformational transitions that follow substrate engagement. The AAA domains of RPT3 and RPT6 bury ~1800 $Å^2$ of interface area with RPN5-RPN7. One side of N-terminal helical repeats in the RPN7 PCI domain, consisting of four helix-connecting loops, makes extensive interaction (~940 $Å^2$) with the RPT6 AAA domain. In contrast, the AAA domains of RPT3-RPT6 contact RPN5 and RPN6, with smaller buried footprints of ~340 $Å^2$ and ~480 $Å^2$, respectively. The interfaces of RPN7-RPT6, RPN6-RPT6 and RPN5-RPT3 may serve as "switches" to allosterically regulate the conformations of the RPT subunits, thus communicating between adjacent subcomplexes during substrate degradation (Greene et al. 2019; Dong et al. 2019).

The CP Subcomplex

The CP has a cylindrical shape, with overall dimensions of 148 × 113 × 75 $Å^3$ (Fig. 1.7). While the archaeal CP is composed of 28 subunits arranged in four homoheptameric rings $\alpha_7\beta_7\beta_7\alpha_7$ coaxially stacked with *D7* symmetry, the eukaryotic CP is assembled from four heteroheptameric rings, $\alpha_{1-7}\beta_{1-7}\beta_{1-7}\alpha_{1-7}$, with only *C2* symmetry. The CP conceals the proteolytic active sites in the inner surface of β-ring. As defined by the characteristics of the P1 cleavage sites of chromogenic reporter groups, the Thr1 residues in β1, β2, and β5 subunits have been ascribed with caspase-like (acidic; peptidylglutamyl-peptide hydrolytic (PGPH)), trypsin-like (basic), and chymotrypsin-like (hydrophobic) activities. However, these specificities are not reflected in the cleavage-site pattern of its substrates that are cleaved at almost arbitrary positions.

In eukaryotic cells, five β-subunits are synthesized as proproteins, which are proteolytically cleaved to become the mature forms during the final steps of proteasome

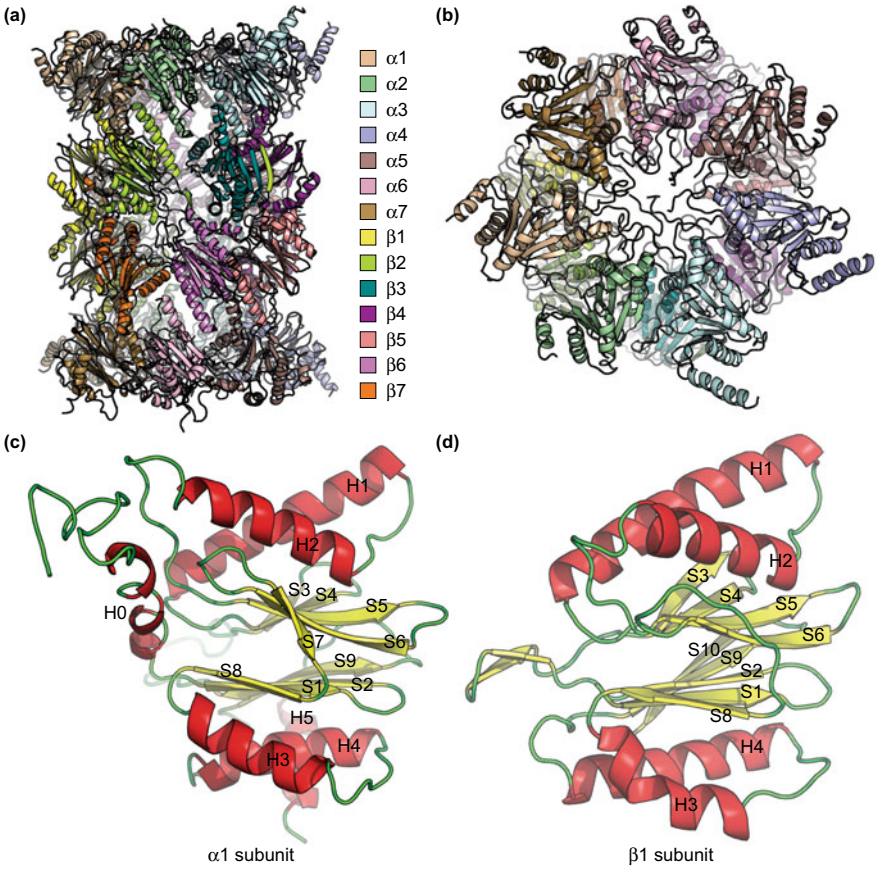


Fig. 1.7 The atomic structure of the human CP (PDB ID: 6MSB) (Dong et al. 2019). **a** The overall structure of the CP from a lateral view. The color codes of the subunits are displayed on the right. **b** The CP structure from a top view. **c** and **d** The atomic models of α1-subunit (**c**) and β1-subunit (**d**) are shown in cartoon representations, which are typically homologous among all the α- and β-type subunits. The names of the secondary structures are labelled

assembly (Chen and Hochstrasser 1995; Seemuller et al. 1995a; Frentzel et al. 1994). The proforms of β1, β2, and β5 subunits are cleaved between Gly(-1) and Thr1, which liberates the active site at residue Thr1. β6 is cleaved between His(-10) and Gln(-9) and β7 is cleaved between Asn(-9) and Thr(-8). β1 and β3 are not further processed and remain in their primary translational forms. All seven α-subunits and seven β-subunits share the common α/β/β/α sandwich-like fold characteristic of Ntn (N-terminal nucleophile)-hydrolase superfamily (Oinonen and Rouvinen 2000). It exhibits two five-stranded antiparallel β-sheets sandwiched by three helices (H3, H4 and H5) on the top and two helices (H1 and H2) on the bottom. Structural differences between the subunits are entailed by the length variations (often by one or two

residues) in the N-terminal and C-terminal regions, in turning loops and in long insertions bridging adjacent secondary structural elements. For instance, the subunits α3 and α4 have longer C-terminal helices (H5) sticking out from the CP surface toward the RP, potentially making direct contact with the base subunits.

The intersubunit interactions mediating the assembly of the CP can be categorized into α-*cis*, β-*cis*, α-*trans*-β, and β-*trans*-β contacts. The α-*cis* contacts are mediated by the N-terminal loop, the helix H0, the strand S7, the loop L, and the loop linking H2 with S5. The β-*cis* contacts are mediated by loop L, the N-terminus of helix H1, strand S7 and the turn connecting helix H3 and strand S8. The α-*trans*-β contacts are mediated by the helix-loop-helix motifs spanning H1 and H2, which interdigitate with the same motifs of two adjacent α-subunits. This mode of intersubunit contacts is conserved in both prokaryotes and eukaryotes (Lowe et al. 1995; Groll et al. 1997; Dong et al. 2019). β-*trans*-β contacts are more specifically formed by the C-terminal arm of β7 intercalating between β1′ and β2′ from adjacent β′-ring, and by the intercalation of β5 C-terminal arm with β3′ and β4′ in a similar configuration.

The CP gate refers to the central axial pore controlled by the N-terminal loops of α-type subunits (Groll et al. 2000). The CP stays in an inactive state in its free 20S form and in the resting state after its assembly with the RP. In this inactive state, the CP gate is blocked by stacking of N-terminal loops of only three α-type subunits, α2, α3 and α4. Particularly, the N-terminal loop of α3 occupies the majority of the space at the CP gate, resembling a linchpin that is stabilized by the N-terminal loops of α2 and α4. Opening of this gate is allosterically controlled by interactions between the CP and RP.

The RP-CP Interface

The RP associates with the CP through a symmetry-mismatched interface (~3600 $Å^2$) between the hexameric AAA-ATPase motor and the heptameric α-ring in the resting state (S_A). All RPT subunits but RPT6 directly contact the α-ring. The large void between the RPT6 and α2 subunits is sided by the N-terminal helical repeats in the PCI domain of RPN6. The helical repeats of RPN6 PCI domain protrudes down to the lateral surface of α2, making ~620-$Å^2$ contact. Similarly, the N-terminal helical repeats of RPN5 PCI domain also make a lateral contact with the α1 subunit via a small footprint area (~50 $Å^2$) (Chen et al. 2016a; Eisele et al. 2018; Zhu et al. 2018; Dong et al. 2019).

In the resting state, the conserved hydrophobic-Tyr-X (HbYX) motifs of RPT3 and RPT5 are docked into the α1-α2 and α5-α6 pockets, respectively (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016) (Fig. 1.8). The terminal oxygen of the penultimate tyrosine in the HbYX motifs (Tyr417 in RPT3 and Tyr438 in RPT5) forms a hydrogen bond with the carbonyl oxygen in the mainchain of the glycine residue in the corresponding α-pocket (Gly20 in α1 and Gly19 in α5). Similar contacts have also been observed between PA26 and archaeal 20S proteasome precursors (Stadtmueller et al. 2010). These structural features explain the notion that RPT3

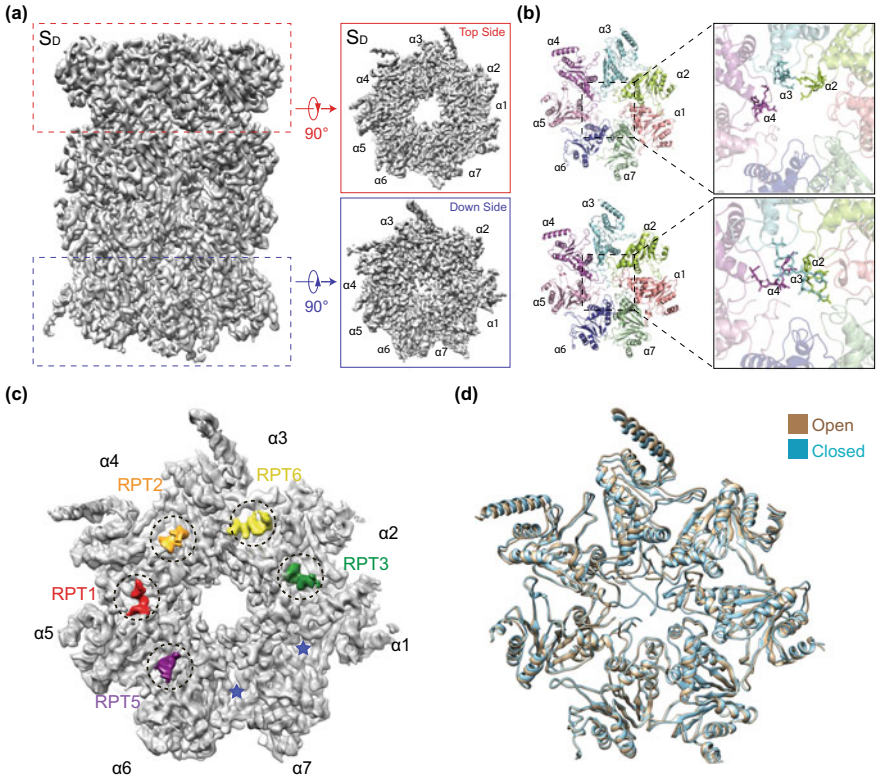


Fig. 1.8 Asymmetric opening of the CP gates in the 26S proteasome (Chen et al. 2016a; Zhu et al. 2018). **a** The cryo-EM density map of the CP in the S_D state viewed from the side (left), the top (upper right), and the bottom (lower right). The RP density is not shown. The upper half of the CP faces the RP in the S_D state and is in the open conformation. **b** The atomic models of the α-rings in the open (upper left) and closed (lower left) states in cartoon representation, viewed from the perspective of the RP-CP interface. The two panels in the right show the central parts of the α-rings in the open (upper right) and closed states (lower right). **c** Cryo-EM density map of the RP-CP interface in the open gate state, in which the local densities of the C-terminal tails of RPT1, RPT2, RPT6, RPT3 and RPT5 are highlighted by different colors for clarity. Two empty α-pockets are marked by asterisks. **d** The atomic models of α-rings in the closed (blue) and open (closed) states are superimposed

and RPT5 are the only two ATPase subunits whose C-terminal HbYX motifs are indispensable for assembly of the human 26S proteasome (Kim and DeMartino 2011). By contrast, in the yeast proteasome, the HbYX motifs of RPT2, RPT3 and RPT5 are inserted into the α-pockets (Tian et al. 2011; Stadtmueller et al. 2010; Ding et al. 2017; Wehmer et al. 2017).

Complex Dynamics of the Proteasome

Common Traits of Proteasome Dynamics and Plasticity

The 26S proteasome was consistently observed to sample multiple coexisting conformers in both *Homo sapiens* and *Saccharomyces cerevisiae* (Unverdorben et al. 2014; Chen et al. 2016a; Luan et al. 2016; Wehmer et al. 2017; de la Pena et al. 2018; Eisele et al. 2018; Zhu et al. 2018; Ding et al. 2019; Dong et al. 2019). Although several conformers of the substrate-free proteasome preserve architectural features of the substrate-bound proteasome, certain conformers and crucial substrate interactions are not accessible in the absence of substrates. This underscores the importance of understanding the conformational landscape of the substrate-engaged proteasome under physiological conditions, which is characterized at large by the seven atomic structures of substrate-bound proteasome as summarized in Table 1.2. Because the conformational state in the RP and its interacting α-ring on one side of a doubly capped proteasome can be different from that of the interacting pair on the other side, the states listed in Table 1.2 refer to the conformations of either the RP-CP subcomplex or the singly capped proteasome. There is currently insufficient evidence showing functional difference between the RP-CP subcomplex in a doubly capped proteasome and the singly capped proteasome (Asano et al. 2015).

There are several common patterns of pronounced conformational changes occurring in the proteasome under distinct biochemical conditions. First, the CP conformation assumes states of either closed or open gate. The human CP structures with a closed gate are virtually identical in all corresponding states including E_{A1}, E_{A2}, E_B, E_{C1}, E_{C2}, S_A, S_B and S_C (Dong et al. 2019). The same is true for the human CP with an open gate in the remaining states. This reflects a bistability of the CP gate that has been also observed for the yeast CP. Second, the lid rotation is generally accompanied with the conformational transitions between different states. Most notably, a prominent lid rotation of 25–40° is commonly observed during state transitions of E_B-to-E_C (substrate-engaged) (Dong et al. 2019), S_A-to-S_B transition (substrate-free) (Chen et al. 2016a; Zhu et al. 2018), C1-to-C3 (tetraubiquitin-bound yeast proteasome) (Ding et al. 2019), and s1-to-s2 (substrate-free yeast proteasome) (Wehmer et al. 2017; Eisele et al. 2018; Ding et al. 2017). Third, the quaternary structural relationship between RPN11 and the OB ring is also approximately bistable. The RPN11 blocks the entrance of the OB ring in states E_{C1}, E_{C2}, E_{D1}, E_{D2}, S_B, S_C, S_{D1}, S_{D2}, S_{D3}, 5D, 5T, 4D, C3-a, C3-b, s2, s3, s4, s5, and s6 (Dong et al. 2019; de la Pena et al. 2018; Eisele et al. 2018; Ding et al. 2019; Chen et al. 2016a; Zhu et al. 2018). By contrast, RPN11 is rotated to the side of the OB ring and leaves a wide entrance to the ATPase channel in all remaining states. Fourth, whenever the CP gate is open in states E_{D1}, E_{D2}, S_{D1}, S_{D2}, S_{D3}, 5T, 4D, C3-a, C3-b, s4, s5, and s6, the conformations of the ATPase ring in these states are always remodeled and translated above the CP to positions that confer a better axial alignment of the ATPase channel with the CP gate as compared to the resting state (S_A). Fifth, the pore-loop staircase is similarly observed in states E_{D1}, 5D and 5T, whereas a different pore-loop staircase

Table 1.2 Comparison of conformational states of the 26S proteasome between *Homo sapiens* and *Saccharomyces cerevisiae* under a variety of biochemical conditions

States	E_{A1}	E_{A2}	E_B	E_{C1}	E_{C2}	E_{D1}	E_{D2}
Functional step	Ubiquitin recognition	Ubiquitin transfer	Deubiquitylation	Initiation of translocation	Priming for CP gate opening	Processive translocation	Processive translocation
Map resolution (Å)	2.8–3.0	3.2	3.3	3.5	3.6	3.3	3.2
Particle percentage	8.4%	6.4%	19.4%	9.0%	5.7%	23.1%	27.9%
Locations of ubiquitin densities	RPN1, RPN10, RPT4/5 CC	RPN1, RPN10, RPN11, RPT4/5 CC	RPN10, RPN11, RPT4/5 CC	RPN11	N/A	N/A	N/A
RPN11 Ins-1 conformation	Large loop	β-hairpin	β-hairpin	β-hairpin	Retracted small loop	Retracted small loop	Retracted small loop
Pore-loop staircase contacting substrate	No substrate inside the ATPase ring, but RPT3 at the top, followed by RPT4/5/1/2		RPT3 at the top, followed by RPT4/5/1/2	RPT6 at the top, followed by RPT3/5/1	RPT6 at the top, followed by RPT3/5/1	RPT1 at the top, followed by RPT2/6/3/4	RPT5 at the top, followed by RPT1/2/6/3
RPT subunits disengaged from pore loop staircase	RPT6	RPT6	RPT6	RPT1 and RPT2	RPT1 and RPT2	RPT5	RPT4
RPT subunits with apo-like state	None	None	RPT6	RPT2	RPT1 and RPT2	RPT5	RPT4
RPT subunits with ADP bound	RPT6 and RPT5	RPT6 and RPT5	RPT2 and RPT4	RPT1 and RPT5	RPT5	RPT4	RPT3
RPT C-tails insertion into α-pockets	RPT3 and RPT5	RPT3 and RPT5	RPT2/3/5	RPT2/3/5/6	RPT2/3/5/6	RPT1/2/3/5/6	RPT1/2/3/5/6
CP gate status	Closed	Closed	Closed	Closed	Closed	Open	Open

(continued)

Table 1.2 (continued)

States	E_{A1}	E_{A2}	E_B	E_{C1}	E_{C2}	E_{D1}	E_{D2}
Conformational states of substrate-free human proteasome (Chen et al. 2016a; Zhu et al. 2018)	**S_A** (3.6 Å) (Similarity in overall structure, but different nucleotide states)		N/A	**S_B** (7.0 Å) (Similarity in the lid-base relationship, but different ATPase structures)	**S_C** (5.8 Å) (Similarity in the lid-base relationship, but different ATPase structures)	**S_{D1}**, **S_{D3}** (4.2, 4.9 Å) (Similarity in the lid-base relationship, RP-CP interface and CP gate, but different ATPase structures)	**S_{D2}** (4.3 Å) (Similarity in the pore-loop staircase, lid-base relationship, RP-CP interface and CP gate)
Conformational states of substrate-bound yeast proteasome (de la Pena et al. 2018)	N/A	N/A	N/A	N/A	**1D*** (4.4 Å) (Similarity in the lid-base relationship, RP-CP interface and CP gate, but different ATPase structures)	**5D**, **5T** (4.7, 4.3 Å) (Similarity in the pore-loop staircase, overall lid-base relationship, RP-CP interface and CP gate)	**4D** (4.2 Å) (Similarity in the pore-loop staircase, overall lid-base relationship, RP-CP interface and CP gate)
Conformational states of tetraubiquitin-bound yeast proteasome (Ding et al. 2019)	**C1-a/b, C2** (3.5, 4.5 Å)		N/A	N/A	N/A	**C3-b** (7.5 Å) (Similarity in the pore-loop staircase, overall lid-base relationship, and CP gate)	**C3-a** (7 Å) (Similarity in the pore-loop staircase, overall lid-base relationship, and CP gate)

(continued)

Table 1.2 (continued)

States	E_{A1}	E_{A2}	E_B	E_{C1}	E_{C2}	E_{D1}	E_{D2}
Conformational states of substrate-free yeast proteasome (Wehmer et al. 2017; Eisele et al. 2018; Ding et al. 2017)	**s1** (4.1 Å)		N/A	**s2** (4.5 Å)	**s3** (5.4 Å), and an **ADP-AlFx-bound state** (4.2 Å)	**s5**, **s6** (4.9, 6.1 Å) (Similarity in the overall lid-base relationship, RP-CP interface and CP gate, but different ATPase conformers)	**s4** (4.5 Å) (Similarity in the pore-loop staircase, overall lid-base relationship, RP-CP interface and CP gate)

Due to their higher resolution, better map quality and relative completeness in defining the energy landscape of the proteasome, the states of the substrate-bound human proteasome in a buffer containing 1 mM ATP and 1 mM ATPγS (Dong et al. 2019) are used as a canonical reference for comparisons with the available states of proteasome under other four distinct categories of biochemical conditions: (1) the substrate-free human proteasome in the presence of 1 mM ATP or ATPγS (Chen et al. 2016a; Zhu et al. 2018), (2) the substrate-bound yeast proteasome with inactivated RPN11 in the presence of 1 mM ATP (de la Pena et al. 2018), (3) the Lys48-linked tetraubiquitin-bound yeast proteasome in the presence of 1 mM ATP (Ding et al. 2019), and (4) the substrate-free yeast proteasome with one of the nucleotides or nucleotide analogs, including the cases of 1 mM ATP, 2 or 4 mM ATPγS, 1 mM ADP, ADP-AlFx, ADP-BeFx, ATP/BeFx, or AMP-PNP (Wehmer et al. 2017; Eisele et al. 2018; Ding et al. 2017). AMP-PNP, adenylyl-imidodiphosphate. AlFx, aluminum fluoride. BeFx, beryllium fluoride. Note that state s6 was found in the *rpt3*-EQ proteasome mutant (Eisele et al. 2018)

is comparably observed in states E_{D2}, S_{D2}, and 4D. Taken together, these common characteristics indicate that the conformational dynamics of the proteasome is evolutionarily conserved to meet its critical functional needs. Despite these shared traits in conformational dynamics of the proteasome holoenzyme, the exact structures of the RP are all somehow different, more or less, among all the states enlisted in Table 1.2, illustrating a structural plasticity of the proteasome to an exceptionally large extent, which is yet to be comparably found in other protein complexes.

Key States of Proteasome in the Act of Substrate Degradation

States E_{A1} and E_{A2} overall resemble the S_A conformation of the substrate-free human proteasome holoenzyme (Dong et al. 2019). Several remarkable features distinguish states E_{A1} and E_{A2} from S_A. Foremost, a ubiquitin density is observed around the T1/T2 site of RPN1 (Shi et al. 2016). Two ubiquitin densities are found at the N-terminal CC domain of RPT4-RPT5 near RPN10. The ubiquitin near RPN11 is attached to the RPT4-RPT5 CC but does not directly contact RPN11 in state E_{A1}. This ubiquitin is bound with RPN11 and detached from the RPT4-RPT5 CC in state E_{A2}. However, no substrate is observed inside the AAA-ATPase motor, indicating that substrate engagement is not yet completed in these states. These suggest that states E_{A1} and E_{A2} represents the steps of initial ubiquitin binding and ubiquitin transfer to RPN11, respectively, presumably before a substrate is fully engaged with the proteasome for deubiquitylation.

State E_B presents a quaternary structure in the act of substrate deubiquitylation and finds no rigorous counterparts among those substrate-free proteasome conformations (Dong et al. 2019). It reveals an interesting conformation that has been absent in all other studies (Chen et al. 2016a; Huang et al. 2016; Schweitzer et al. 2016; de la Pena et al. 2018; Zhu et al. 2018). Although the overall RP-CP relationship is closer to states S_A and E_A, the lid swings outward away from the OB ring relative to state E_A, toward an opposite direction against that seen in the E_B-to-E_C transition. This results in a wider access to the AAA-ATPase ring. Notably, the isopeptide bond linking substrate lysine to the C-terminal glycine of RPN11-bound ubiquitin has been observed in the vicinity to the zinc ion bound at the catalytically active site of the DUB RPN11. This DUB-substrate interaction is achieved through extensive quaternary rearrangements also involving RPT5 and RPN8. The lysine residue linking to the RPN11-bound ubiquitin via the isopeptide bond is located above the OB ring and is about 15 Å away from the central pore of the ATPase ring, where the substrate density is observed all the way from the ubiquitin-conjugated lysine down to the end of the central pore of the ATPase ring.

States E_{C1} and E_{C2} present two successive snapshots that capture the initiation steps of substrate translocation (Dong et al. 2019). While the RPN11-bound ubiquitin remains in state E_{C1}, its isopeptide bond linking the substrate appears to have been broken. In state E_{C2}, this ubiquitin is released from RPN11. As the overall lid-base relationship resembles the substrate-free state S_C, the lid is rotated ~40° clockwise relative to states E_A and E_B. Although the ATPase conformation is nearly invariant in

states E_{C1} and E_{C2}, the ATPase ring exhibits a rigid-body translation of ~5 Å above the CP, with a small rotation in the lid. The substrate inside the AAA-ATPase ring is presumably advanced toward the CP relative to state E_B. The two states exhibit similar features in the RP-CP interface with a closed CP gate.

States E_{D1} and E_{D2} present two sequential conformations in which substrate-pore loop interactions are consistent with processive substrate translocation (Dong et al. 2019). The substrate is presumably moved toward the CP relative to states E_{C1} and E_{C2}. No ubiquitin densities are found on RPN11. The overall lid-base relationship, the RP-CP interface and the open state of CP gate are compatible with the substrate-free states S_{D1}, S_{D2} and S_{D3} of the human proteasome (Chen et al. 2016a; Zhu et al. 2018), and the substrate-engaged states 5T, 4D (de la Pena et al. 2018), the tetraubiquitin-bound states C3-a, C3-b (Ding et al. 2019), and substrate-free states s4, s5 and s6 of the yeast proteasome (Wehmer et al. 2017; Eisele et al. 2018). It is noteworthy that the pore-loop staircase architecture of states 5D, 5T, C3-b and s6 closely resembles that of state E_{D1}, whereas the pore-loop topologies of states S_{D2}, 4D, C3-a and s4 match state E_{D2}.

Allosteric Regulation of Proteasome Dynamics

Ligand binding and substrate interactions substantially modify the energy landscape of the proteasome and thus regulate both the distribution of coexisting conformers and the key features of these conformers (Fig. 1.9). In the presence of ATP/Mg^{2+}, the substrate-free human proteasome exhibits four conformational states S_A, S_B, S_C and S_D in the particle populations at 76.1%, 10.2%, 5.8% and 7.9%, respectively. In the presence of ATPγS (that completely replaces ATP), the expanded states S_A, S_B, S_C, S_{D1}, S_{D2} and S_{D3} represent 51.8%, 3.5%, 5.3%, 14.9%, 17.0% and 7.5% of the total particles, respectively (Zhu et al. 2018). The particle population of the human proteasome in an open-gate state is about five times higher in the presence of ATPγS than ATP. Indeed, the peptide-hydrolyzing activity of the proteasome has been observed to be substantially stimulated by ATPγS (Li and Demartino 2009). In contrast to the shift of conformational equilibria toward the open-gate states, binding of the small-molecule inhibitor Oprozomib to the β5 subunit in the CP stabilizes the resting state (S_A) by substantially narrowing the ground funnel on the energy landscape and inhibiting the human proteasome from sampling those states with an open CP gate (Haselbach et al. 2017).

Structural and biochemical studies on the substrate-free yeast proteasome also agree with this ligand-dependent regulatory picture in general. Binding of ATPγS or other nucleotide analogs, such as AMP-PNP, ADP-AlF_x and ADP-BeF_x, substantially induces ATPase conformations resembling those substrate-engaged states and shifts the conformational landscape of the proteasome from the resting state (s1) toward the open-gate states (s4, s5 and s6) (Wehmer et al. 2017; Eisele et al. 2018; Ding et al. 2017). Similar allosteric effects have also been observed for the Lys48-linked tetraubiquitin-bound yeast proteasome (Ding et al. 2019). In this case, the conformational equilibria of the proteasome are driven toward two open-gate states

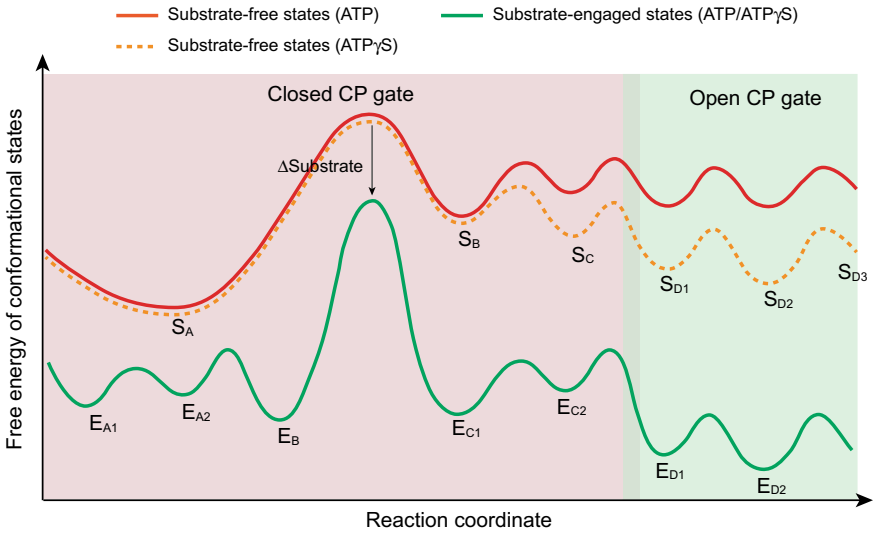


Fig. 1.9 Conceptual schematic illustrating the regulation of the energy landscape of the 26S proteasome by substrates and nucleotides. The replacement of ATP with slowly hydrolyzed ATPγS enhanced the conformational states corresponding to an open CP gate. The engagement of a substrate with the proteasome further stabilizes these states and induces more sub-states corresponding to a closed CP gate

(C3-a and C3-b), which resemble the substrate-engaged human states E_{D2} and E_{D1}, respectively (Dong et al. 2019). Notably, the resting state no longer dominates the energy landscape in some conditions. For example, 2 mM ATPγS significantly expands the particle population of state s3, s4 and s5, whereas 4 mM ATPγS even further promotes the distribution of state s3 and s4, allowing the two states to dominate the energy landscape of the yeast proteasome (Eisele et al. 2018).

Because the resolution of most yeast proteasome reconstructions was insufficient to define the nucleotide states in the proteasomal ATPases, a genetic approach has been developed to evaluate how individual ATP-binding events influence the conformational landscape of the proteasome (Eisele et al. 2018). The conserved Walker B glutamate was mutated to glutamine in each ATPase subunit in yeast, which was codenamed *rpt*-EQ. While mutations in three ATPases caused cytotoxic effects, mutant proteasomes from *rpt2*-EQ, *rpt*3-EQ, and *rpt*6-EQ enabled cryo-EM analysis that showed a consistent stimulation of states s4, s5 and even discovery of a new conformational state s6 of the yeast proteasome. While states s4 and s5 were observed in *rpt2*-EQ and *rpt*6-EQ proteasomes, state s6 was exclusively observed in *rpt*3-EQ proteasome, in which the ATPase conformation resembles state E_{D1}.

In the presence of both ATP/Mg^{2+} and ATPγS/Mg^{2+}, the substrate-engaged human proteasome shows seven conformers E_{A1}, E_{A2}, E_B, E_{C1}, E_{C2}, E_{D1} and E_{D2} in the particle populations at 8.4%, 6.4%, 19.4%, 9.0%, 5.7%, 23.1% and 27.9%, respectively (Dong et al. 2019). The open-gate states E_{D1} and E_{D2} dominate the energy

landscape, in stark contrast to the conformational distributions of substrate-free proteasome. Because all these seven states have been reconstructed at a resolution no worse than 3.6 Å, these human proteasome structures afforded unambiguous determination of the nucleotide states of the ATPase, allowing the elucidation of how the nucleotide hydrolysis and exchange trigger state transitions of the proteasome (Dong et al. 2019). Similar allosteric effects were also noted for the substrate-bound RPN11-inactivated yeast proteasome (de la Pena et al. 2018). These structural data explain why substrate engagement considerably stimulates the ATPase activity and serve as a foundation to further understand how ubiquitin-like (UBL) proteins stimulate proteasome activation (Kim and Goldberg 2018; Collins and Goldberg 2020).

In vivo experimental studies using cryo-electron tomography (cryo-ET) and subtomogram classification suggest that the 26S proteasome mostly stays at the resting state in intact hippocampal neurons in the absence of proteotoxic stress, and exhibits several alternative conformations (Asano et al. 2015; Guo et al. 2018). Although the subtomogram averaging from cryo-ET reconstructions has a low resolution compared to single-particle reconstruction, the large RP rotation between the resting state (S_A) and the substrate-degrading states (E_D) is discernible at a nanometer resolution. Subtomogram classification has been conducted on the proteasome within intact neurons showing poly-Gly-Ala (poly-GA) aggregates, which results from aberrant translation of an expanded GGGCCC repeat in the *C9orf72* gene. Mutation in *C9orf72* is the most common cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (Edbauer and Haass 2016; Freibaum and Taylor 2017; Gendron et al. 2013; Lin et al. 2017). The poly-GA aggregates in neurons densely pack into twisted ribbons and recruit a large number of 26S proteasomes that exist in either the substrate-degrading states (E_D) or the resting state (S_A) (Guo et al. 2018). The cryo-ET studies provide strong evidence supporting that the in vivo proteasome conformational dynamics regulated by the endogenous substrates are reflected in those analyzed in vitro at the atomic level.

Activation Dynamics of the CP

By studying the *Thermoplasma acidophilum* CP, which provides a lower structural complexity, the NMR quantification of structural dynamics of activator binding and specifically labeled residues (Sprangers and Kay 2007) elucidated the mechanism of CP gating by the unstructured N-termini of α-subunits in the archaeal proteasomes. It was observed that on average two of the chains move through the CP gate to the proteolytic chamber, and block the passage for translocating protein substrates (Religa et al. 2010). The NMR studies of the CP were conducted by methyl transverse relaxation optimized spectroscopy, which takes advantage of deuterated proteins and selectively labelling methyl groups of isoleucine, leucine, methionine and valine (Kay 2011).

How the CP gate is regulated by its activators has been informed by numerous structural studies. Crystal structure of the human PA28α/REGα homolog reveals a toroidal heptamer architecture. The CP-interacting C-terminal residues and

internal activation loop exhibit sevenfold symmetry (Knowlton et al. 1997). The 20-Å pore inside this heptamer was initially suggested to be a substrate translocation channel. However, it was later found that this pore is occupied in the PA26 homolog of *Trypanosoma brucei* distantly related to PA28 (Forster et al. 2005). Crystal structures of PA26 in complex with the CPs in *Saccharomyces cerevisiae* (Whitby et al. 2000; Forster et al. 2003) and *Thermoplasma acidophilum* (Forster et al. 2005) indicate that the C-termini of PA26 are docked into the α-pockets, whereas the activation loops trigger the opening of the CP gate by repositioning the Pro17 turn of the CP in a sevenfold symmetric fashion.

Crystal structures of a Blm10-CP complex exhibit an architecture distinct from that of PA28. The single-chain ~250-kDa Blm10 activator binds the α-ring and resembles the shape of a turban (Iwanczyk et al. 2006; Sadre-Bazzaz et al. 2010; Schmidt et al. 2005). Interestingly, Blm10 docking onto the CP induces a disordered conformation of the CP gate. There appears to be only limited access to the turban-like architecture of Blm10 above the CP gate, consistent with a lower level of peptidase activities in the presence of Blm10 compared to PA26 (Iwanczyk et al. 2006). One C-terminal tail of PA200 docks the α5-α6 pocket, with the three C-terminal residues closely resembling the C-termini structure of PA26, forming comparable mainchain hydrogen bonds and electrostatic interactions (salt bridge) with the surface-exposed lysine in the α6 subunit. This is insufficient to open the CP gate, because other α-pockets are not occupied and conserved Blm10 residues sterically impedes the gate opening. The structures of the PA28 and Blm10/PA200 in complex with the CP point to the same binding mechanism of the CP activators in general. In both cases, a salt bridge is formed between the C-terminal carboxylate of the activator and the surface-exposed lysine of the CP (Forster et al. 2005; Dange et al. 2011; Sadre-Bazzaz et al. 2010) and the Pro17 turns of the CP is repositioned, resembling their configuration in the PA26-CP complexes. These features underscore the functional importance of the penultimate tyrosine (or phenylalanine) in the C-termini of the CP activators (Smith et al. 2007). This mechanism of CP gating has been further supported by crystal structures of the PA26 mutants binding the archaeal CP (Stadtmueller et al. 2010; Yu et al. 2010), a cryo-EM structure of the CP in complex with the PAN C-terminal peptides (Rabl et al. 2008), and high-resolution cryo-EM structures of the human 26S proteasome (Chen et al. 2016a; Zhu et al. 2018; Dong et al. 2019).

The structures of the PA28 and PA200 complexes do not intuitively clarify their elusive biological function (Rechsteiner and Hill 2005). The expression of PA28α and PA28β can be induced by interferon-γ, implicating a potential role in T-cell antigen presentation mediated by major histocompatibility complex class I (MHC-I) (Sijts et al. 2002). However, the mechanism underlying this process remains elusive, because many species that do not seem to encode MHC-I also express a PA28 homolog (Sijts and Kloetzel 2011). PA28γ/REGγ was observed to mediate ATP-independent and ubiquitin-independent degradation of some natively unstructured transcription factors (Chen et al. 2007; Li et al. 2007a). PA200 has been proposed to play a role in maintenance of mitochondrial inheritance, in DNA repair and in spermatogenesis. The abundance of the PA200-CP complexes was increased when either activated or disrupted CP prevails (Savulescu and Glickman 2011). Both PA28

and PA200 can participate in the assembly of hybrid proteasomes, in which different types of CP regulators bind to opposite sides of the same CP (Cascio et al. 2002; Kopp et al. 2001; Tanahashi et al. 2000). Such hybrid proteasomes with the PA28 or PA200 at one side and the RP at the other side have indeed been observed on the endogenously purified eukaryotic proteasome by cryo-EM in my laboratory (unpublished data). The hybrid proteasome is expected to possess distinct gating dynamics on opposite sides of the same CP. Interestingly, the gate conformations at the opposite sides of the archaeal CP appear to be coupled, which might allosterically regulate the assembly of hybrid proteasomes (Yu et al. 2020). Structural and functional studies of the hybrid proteasomes are anticipated to clarify their roles in regulating cellular processes.

In the 26S proteasome, the conformation of the CP gate is controlled by the RP-CP interface in an ATP-dependent fashion. The intermediate states of the RP-CP interface are characterized in the substrate-engaged proteasome. Perhaps the most interesting observation at this critical interface is the insertion of one additional C-terminal tail of RPT subunits into one α-pocket at a time, when the proteasome navigates from state E_A, through states E_B and E_C to state E_D. In the resting state (S_A) or substrate-accepting states (E_{A1} and E_{A2}), RPT3 and RPT5 have already docked their C-terminal HbYX motifs into two α-pockets (α1-α2 and α5-α6). State transition to E_B is accompanied with RPT2 HbYX motif insertion into the α3-α4 pocket, followed by RPT6 C-terminus insertion into the α2-α3 pocket in State E_C and RPT1 HbYX motif insertion into the α4-α5 pocket in State E_D (Fig. 1.10) (Dong et al. 2019). Particularly, the RP-CP interaction modes of substrate-engaged human states E_C and E_D were consistently found in those of substrate-free yeast states s3 and s4-s6, respectively (Eisele et al. 2018).

The principal gate-blocking N-terminal loops are from the α2 and α4 subunits (Groll et al. 2000). The α3 N-terminal loop behaves as a lynchpin of the CP gate overlaid with the N-terminal loops of α2 and α4 (Tian et al. 2011). Destabilization of the α3 lynchpin is necessarily but insufficiently controlled by RPT2. By comparison, the yeast RP-CP interaction in the resting state (s1) exhibits the same mode observed in state E_B of the human proteasome, in which three HbYX motifs from RPT2, RPT3 and RPT5 are inserted in the α-pockets (Wehmer et al. 2017; Dong et al. 2019). In contrast to the sequential insertion of RPT C-terminal tails in the human proteasome during the process of CP activation, the yeast CP apparently requires fewer intermediate steps in full activation (Eisele et al. 2018). It is also notable that once four RPT C-terminal tails docked into the α-pockets, the CP gate starts to show partial opening, as captured in the substrate-bound state 5D (de la Pena et al. 2018), states E_{C1} and E_{C2} (Dong et al. 2019) and substrate-free state s3 (Wehmer et al. 2017).

The insertion of five RPT C-terminal tails also demands the repositioning of the ATPase ring above the CP, as well as necessary RPT conformational rearrangements to satisfy the geometrical requirement. In states E_D or S_D, the N-terminal loops of α2, α3 and α4 are rotated ~90° to approximately align along the heptameric axis to open the CP gate. However, the secondary structural elements surrounding the gate-blocking loops in the α-ring remain largely unchanged with only very subtle

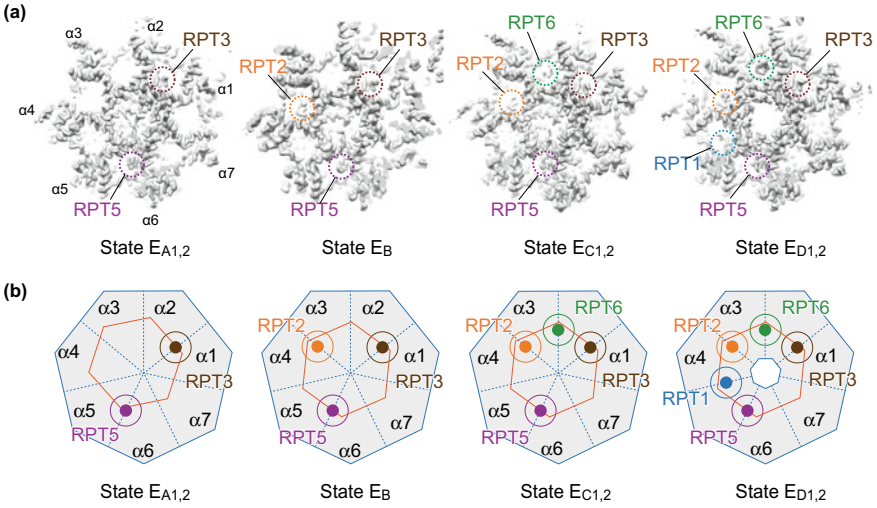


Fig. 1.10 Gating mechanism of the CP in the human proteasome (Chen et al. 2016a; Zhu et al. 2018; Dong et al. 2019). **a** Comparison of the cryo-EM densities of the RPT C-termini binding the α-pockets of the CP in different conformational states of the human proteasome. The cryo-EM reconstructions of the RP-CP interfaces are shown in grey surface representation. The colored dashed circles highlight the densities of the RPT C-terminal tails. **b** A schematic diagram summarizes the RPT C-terminal interaction with the α-pockets and the state of the CP gate in all states shown in (**a**). The ATPase ring is represented as a hexagon, the CP as a heptagon and the RPT C-terminal interaction as a colored sphere

movements relative those in other states; indeed, the overall outer diameter of the opened α-ring is dilated by only ~3 Å (Zhu et al. 2018; Dong et al. 2019). A similar allosteric effect has also been observed in the open gate of yeast CP in complex with the PA26/11S regulator from *Trypanosoma brucei* (Whitby et al. 2000; Stadtmueller et al. 2010). Importantly, in both the RP-activated human CP and the PA26-activated yeast CP, the same α-pockets (α6-α7 and α7-α1) are the only two that are empty, indicating a highly conserved mechanism of CP gating regulation by its activators.

The CP is not only externally regulated by the RP, but is also internally affected by the interactions at the proteolytic sites in the β-type subunits (Haselbach et al. 2017; Osmulski et al. 2009). Studies of the yeast CP in the absence of RP using atomic force microscopy have suggested that the CP gating is either stimulated or suppressed by the proteolytic active-site engagement with a substrate or small-molecule inhibitors such as Bortezomib and Epoxomicin (Osmulski et al. 2009). The CP gate opening appears to be stimulated by a tetrahedral transition state at the proteolytic active sites. Notably, such a short-range allosteric effect on the CP gate can be further transmitted to the conformational state of RP through the RP-CP interface in a long-range fashion (Haselbach et al. 2017).

Quaternary Allosteric Regulation

The activity of proteasomal ATPase motor is regulated by the lid subunits, RPN1 and RPN2 in an asymmetric fashion (Greene et al. 2019). While in states E_{D1} and E_{D2}, RPN1 and RPN2 form a dimer by a long-range association through a long turning loop from RPN2, such a quaternary architecture was less clearly observed in other states ($E_{A\text{-}C}$) (Dong et al. 2019). During state transition from E_A to E_B, RPN1 undergoes a prominent rotation. In states E_{C1} and E_{C2}, the RPN1 density is blurred, indicating strong motions potentially disrupting the RPN1-RPN2 associations observed in states E_{D1} and E_{D2}. Thus, the variation of RPN1 motion appears to be highly coordinated with the ATP hydrolytic cycle coupled with conformational remodeling of the ATPase motor during substrate translocation.

The CC domains of RPT4-RPT5 also seem to coordinate substrate translocation by its interaction with the lid subunit. In states E_{C1}, E_{C2}, E_{D1} and E_{D2}, the RPT4-RPT5 CC domain interacts with either RPN9 or RPN10 in different contacts. By contrast, it resides several nanometer away from any RPN subunit and recruits peptide-proximal ubiquitin in state E_A and transfers it to RPN11 for deubiquitylation in state E_B. Thus, the lid-base interaction regulates substrate interactions with the proteasome through a combination of short-range and long-range allosteric regulations, which are supported by a recent biochemical study (Snoberger et al. 2018). The conformational regulation between the lid and base or CP is similarly observed in the absence of substrates (Zhu et al. 2018).

The lateral RP-CP interface, particularly between the RPN6 PCI domain and the α2 subunit, also demonstrates prominent fluctuations that interconvert between open and closed configurations. This interfacial instability is more prominently observed in the states corresponding to an open CP gate or in the presence of substrates, indicating an ancillary role of this interface in the regulation of the conformational changes of the axial substrate-translocation pathway (Zhu et al. 2018; Dong et al. 2019). Indeed, the interaction of RPT6 with the α2 subunit is regulated by phosphorylation on RPT6, which stimulates proteasome activity (Satoh et al. 2001; Asai et al. 2009)

Summary of Proteasome Dynamics

In this section, complex dynamics of the proteasome in the act of substrate processing at the atomic level are compared in detail with structural characterization of proteasome dynamics at the near-atomic or pseudo-atomic level under various in vitro biochemical conditions. This retrospective comparison on 29 conformational states of the proteasome not only reveals the common traits of proteasome dynamics, but also makes clear several outstanding revelations that were relatively weak or less obvious in single studies, as summarized in the following bullet points.

- The resting state is least active in its CP in the presence of ATP. Replacing ATP with other nucleotides or nucleotide analogs, the presence of ubiquitin and UBL

proteins, or the engagement of substrates generally re-equilibrates the proteasome conformations toward the open-CP states and stimulates proteasome activation.

- Both the RPN11-OB interface and the CP gate controlling the substrate entrance are approximately bistable. Their state transitions are desynchronized to coordinate the timing between deubiquitylation and translocation events.
- The lid rocking is generally observed to regulate both the ATPase and CP conformations. It becomes increasingly apparent that the lid interactions with the base and CP favor two sets of ATPase conformations characterized in two sets of comparable states [E_{D1}, 5D/5T, C3-b, s6] and [E_{D2}, S_{D2}, 4D, C3-a, s4]. (See below: section *Operating Principles of Proteasomal AAA-ATPase Motor*, for more details).
- Substrate engagement stabilizes the states with the open CP gate, and via its interactions with the proteasome at multiple sites along the substrate-translocation pathway, coordinates the actions of ubiquitin recognition, deubiquitylation, initiation of unfolding and processive translocation through the ATPase ring and the CP gate. Such a multi-enzymatic coordination is allosterically assisted by the lid subcomplex interacting with the base and CP at multiple peripheral sites.
- The dynamic CP gate is regulated both externally and internally via allosteric effects. This enables the CP gating to be highly responsive to inter-subcomplex coordination for processive substrate degradation without generating proteolytic errors.
- The allosteric regulation between RP and CP is bidirectional via both short-range and long-range pathways. While the interactions of ubiquitin receptors with ubiquitin or UBLs can allosterically impact the CP state, the interaction of the CP proteolytic active sites with substrates or inhibitors can also influence the RP state. Once the RP and CP are assembled into a holoenzyme, the conformational dynamics of all subunits are allosterically, coherently, and globally coupled. No single subunit acts alone without triggering coherent allosteric effects within the holoenzyme.

Life Cycle of the Proteasome

Proteasome Biogenesis

Constitutive Proteasome Expression

The cellular abundance of the proteasome is dynamically regulated by a genetic circuit of negative feedback at the transcriptional level (Motosugi and Murata 2019) (Fig. 1.11). Specific transcription factors regulating expression of the proteasome subunits have been identified (Budenholzer et al. 2017; Rousseau and Bertolotti 2018). In *Saccharomyces cerevisiae*, a transcription factor Rpn4 control coordinated gene expression of proteasome subunits (Wang et al. 2008). Rpn4 is a zinc-finger protein and binds a conserved sequence motif (5′-GGTGGCAAA-3′) known

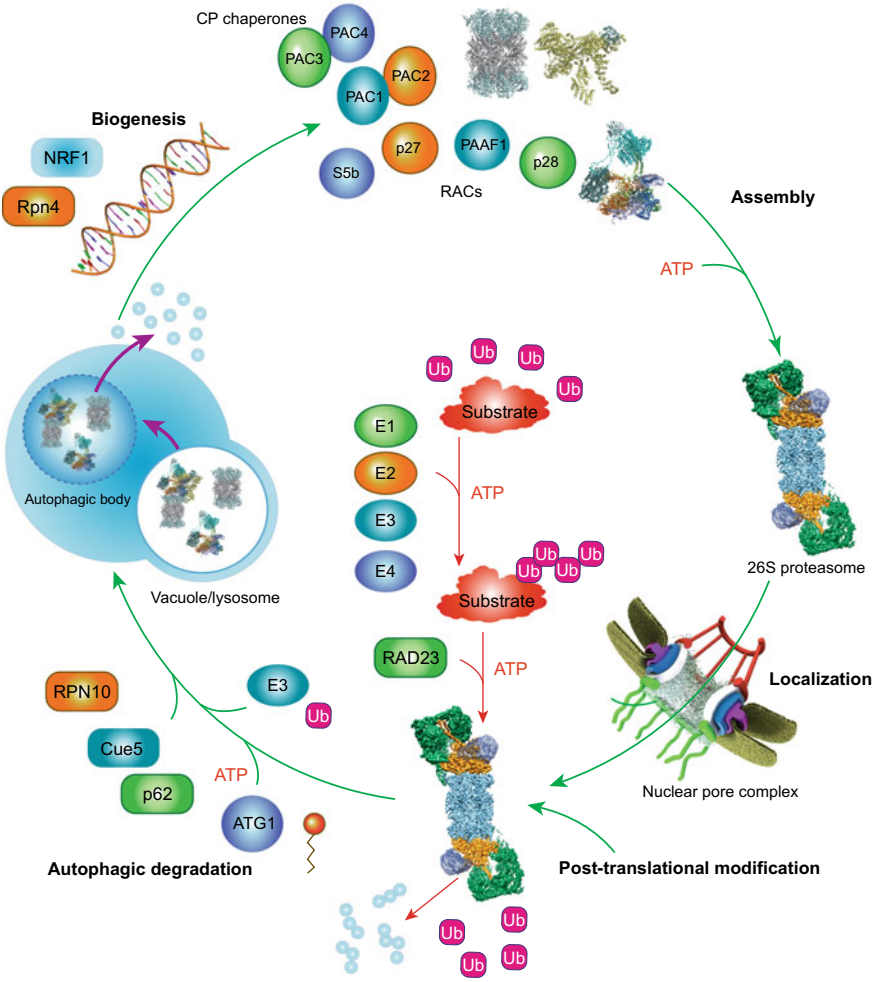


Fig. 1.11 Life cycle of the 26S proteasome. Biosynthesis of proteasome subunits from individual amino acids is regulated by transcription factors such as NRF1 in mammals, NAC53/NAC78 in plants and Rpn4 in yeast. With assistance of CP and RP assembly chaperones, the expressed subunits assemble into the mature form of the 26S proteasome in a coordinated manner. Proteasomes are localized in either the nucleus or cytosol, where their activities can be regulated by numerous proteasome-associated proteins and by post-translational modifications. Extrinsic and intrinsic ubiquitin receptors recognize and deliver polyubiquitylated substrates to the proteasome for degradation. Damaged proteasomes are degraded in the lysosome or vacuole via autophagy, which is mediated by signals from the nutrient-responsive ATG1 kinase, subunit ubiquitylation, and several autophagy receptors, including p62/SQSTM1 in mammals, RPN10 in plants, and Cue5 in yeast

as proteasome-associated control element (PACE) in the promoter region of genes expressing proteasomal subunits and assembly chaperones (Mannhaupt et al. 1999; Shirozu et al. 2015). Rpn4 is rapidly degraded by the proteasome and is thus short-lived ($t_{1/2} \approx 2$ min). Through such a negative feedback circuit, Rpn4 promotes proteasome expression whenever the proteasome function is compromised (Xie and Varshavsky 2001). Loss of *RPN4* gene or PACE sequences compromises proteasome activity and cell viability under various stress conditions such as oxidative DNA damage (Wang et al. 2008). Rpn4 abundance is also regulated by many stress-inducible transcription factors, such as Hsf1, Pdr1, Pdr3, and Yap1 (Owsianik et al. 2002; Ma and Liu 2010). Increasing proteasome expression might be a common mechanism for cells to adapt to diverse challenging conditions (Hahn et al. 2006).

Although mammalian cells lack of any Rpn4 orthologs, several transcription factors regulating the mammalian proteasome expression have been identified. One of such transcription factors is nuclear factor erythroid-derived 2-related factor 1 (NFE2L1) also named NRF1 (Radhakrishnan et al. 2010). NRF1 promotes expression of all genes encoding proteasome subunits and their assembly chaperones in the presence of proteasomal dysfunction, leading to de novo proteasome synthesis. In *Arabidopsis thaliana*, the transcription factors NAM/ATAF1/CUC2 78 (NAC78) and NAC53 were proposed to play a similar role as Rpn4 and NRF1 in upregulating proteasome synthesis that help plants survive proteotoxic stress (Gladman et al. 2016).

In contrast to NRF1, several transcription factors specifically regulate the induction of a portion of proteasome subunits in mammalian cells. Nuclear transcription factor Y (NF-Y) is a complex consisting of three proteins, NF-YA, NF-YB, and NF-YC. NF-Y recognizes the CCAAT motif in the promoter region of certain proteasomal genes and upregulates cellular proteasome activity. The CCAAT motif is found in the genes of six CP subunits (α2, α5, α7, β3, β4, and β6), five RP subunits (RPT1, RPT5, RPT6, RPN10, and RPN11) and one RP assembly chaperone (p28) (Xu et al. 2012). Other transcription factors, such as the forkhead box protein O4 (FOXO4) and the signal transducer and activator of transcription 3 (STAT3), exhibit seemingly greater specificity in the regulation of proteasome subunit expression. FOXO4 specifically regulates RPN6 expression and maintains the proteasome activity at a higher level in human embryonic stem cells (Vilchez et al. 2012; Webb and Brunet 2014). In the JAK/STAT pathway, JAK phosphorylation upon cytokine signaling activates STAT3, which upregulates *PSMB5* (β5 subunit) expression through epidermal growth factor (EGF) induction (Vangala et al. 2014).

Proteasome abundance is also regulated at the post-transcriptional or translational level (Rousseau and Bertolotti 2016; Kors et al. 2019; Motosugi and Murata 2019). In yeast, the mitogen-activated protein kinase 1 (Mpk1) regulates expression of the proteasome subunits and RP assembly chaperones (RACs) under stress conditions or in response to rapamycin and tunicamycin. However, mRNA levels are comparable between wild-type and Mpk1-deficient cells, suggesting that Mpk1-mediated regulation of the proteasome occurs at the translational level. Moreover, the translation of RPT5 mRNA can be suppressed by haploinsufficiency of ribosomal genes, which primarily causes Diamond-Blackfan anemia (Khajuria et al. 2018). In hematopoietic

stem and progenitor cells, translation of proteasome mRNA is selectively inhibited by the storage of ribosomes, thus impeding erythroid lineage commitment.

The growth-controlling kinase mTORC1 (Mammalian or mechanistic target of rapamycin complex 1) is activated in response to growth factors or increased nutrient levels. It activates protein translation and lipid synthesis to promote cell growth and proliferation (Dibble and Manning 2013). The mTORC1 pathway has been found to control proteasome abundance at multiple levels (Zhang et al. 2014; Zhang and Manning 2015; Rousseau and Bertolotti 2016; Zhao et al. 2015). The inhibition of mTORC1 can induce ERK5, the human orthologs of Mpk1, which upregulates the RACs and proteasome subunits (Rousseau and Bertolotti 2016). Stimulated by growth factor signaling, mTORC1 can activate sterol regulatory element binding protein (SREBP-1) that regulates expression of lipogenic genes (Ricoult and Manning 2013). The mTORC1-activated SREBP-1 is related to NRF1 expression that regulates proteasome expression (Zhang et al. 2014). On the other hand, mTORC1 also phosphorylates and inactivates ATG proteins required for autophagy induction, such as ATG1 and ATG13 (Hosokawa et al. 2009; Dibble and Manning 2013). Since the proteasome is destructed by autophagy (Marshall and Vierstra 2015; Marshall et al. 2015; Marshall et al. 2016; Cohen-Kaplan et al. 2016; Waite et al. 2016; Nemec et al. 2017), mTORC1 regulates both the proteasome biosynthesis and breakdown in response to nutrient conditions.

Immunoproteasome Expression

Proteasomal subtypes are differentially expressed in response to certain cellular conditions to meet specific functional needs (Motosugi and Murata 2019; Ferrington and Gregerson 2012). Immune cells constitutively express the immunoproteasomes at high levels. The immunoproteasome primarily processes antigens for their presentation in the MHC-I pathway by degrading the antigens into short polypeptides. The transporter associated with antigen processing (TAP) complex, a TAP1-TAP2 heterodimer, then transports the antigen polypeptides to the ER, where the MHC-I molecules recognize the polypeptides for antigen presentation on the cell surface (Leone et al. 2013). The three catalytic subunits β1, β2, and β5 are replaced by subtype subunit β1i, β2i, and β5i, respectively, in the immunoproteasome (Kloetzel 2001). While *PSMB10* (β2i) is located outside the MHC locus, *PSMB9* (β1i) and *PSMB8* (β5i) resides in the MHC-II region next to the genes encoding TAP1-TAP2.

The expressions of β1i, β2i, and β5i are induced by interferon-γ in response to oxidative stress or inflammatory stimuli (Aki et al. 1994; Hallermalm et al. 2001; Hussong et al. 2010; Hisamatsu et al. 1996). Interferon-γ induces protein expression involved in antigen presentation, including the proteasome activator PA28α/β, and TAP1/TAP2 (Ma et al. 1997; Realini et al. 1994; Ahn et al. 1995). β1i expression requires the GC boxes and the transcription factor SP1 (Wright et al. 1995). The genes encoding β1i and TAP1 share a bidirectional promoter that lacks any TATA box and includes several GC boxes, interferon consensus sequence 2 and γ-interferon-activated sequence sites, which are recognized by the interferon regulatory factor 1

(IRF1) and STAT1 (Wright et al. 1995; Chatterjee-Kishore et al. 2000). β1i expression is induced by IRF1 and STAT1 under interferon-γ stimulation. The nuclear factor-κB (NF-κB), a family of transcription factors, is required for the tumor necrosis factor-α (TNF-α) dependent induction (Wright et al. 1995). The β2i promoter contains NF-κB, IRF1 and SP1 consensus sequences that lack the CAAT or TATA boxes (Cruz et al. 1997; Hayashi et al. 1997). The β5i promoter contains an NF-κB consensus sequence, a GC-rich region and a TATA box (Zanelli et al. 1993). In neurons, expression of β1i and β5i is induced by the transcription factor Zif268 (James et al. 2006).

Thymoproteasome Expression

Thymic cortex specifically expresses the thymoproteasome, another proteasome subtype that plays an important role in the selection of $CD8^+$ T cells (Murata et al. 2007; Murata et al. 2018). In the thymic cortex and medulla, immature thymocytes are subject to positive and negative selection, respectively. In contrast to the immunoproteasomes expressed in medullary thymic epithelial cells (mTECs), the thymoproteasomes is assembled from the subtype subunit β5t in replacement of β5 and is specifically expressed in cortical thymic epithelial cells (cTECs). The *PSMB11* gene encoding β5t is adjacent to *PSMB5* encoding β5. The gene product is encoded by a single exon in both the human and mouse genomes. The forkhead transcription factor FOXN1 promotes both TEC lineage specification and β5t expression (Romano et al. 2013; Uddin et al. 2017). The β5t promoter contains the conserved FOXN1-binding sequence 5′-ACGC-3′, in which a single mutation impairs $CD8^+$ T cell production and β5t expression. Paradoxically, because FOXN1 is present in mTECs that do not express β5t at any detectable level, FOXN1 alone does not sufficiently regulate β5t expression, suggesting that other unknown cellular factors might contribute to the induction of β5t expression.

Proteasome Assembly

CP Assembly Chaperones

The assembly of the eukaryotic CP is more complex than their prokaryotic counterparts and requires a set of extrinsic and intrinsic chaperones (Table 1.1) (Murata et al. 2009; Tanaka 2009; Rousseau and Bertolotti 2018; Budenholzer et al. 2017). The complex process of CP assembly can be understood as three consecutive steps in eukaryotes (Fig. 1.12a) (Murata et al. 2009; Tomko and Hochstrasser 2013; Rousseau and Bertolotti 2018): (1) α-ring formation, (2) β-ring formation and (3) half-CP dimerization and proteasome maturation. These intermediate assembling steps are assisted by five proteasome chaperones named proteasome assembling chaperone-1 (PAC1)-PAC4 and proteasome maturation protein (POMP) in human (Witt et al. 2000; Murata et al. 2009). The yeast orthologs of PAC1-PAC4 and POMP are known

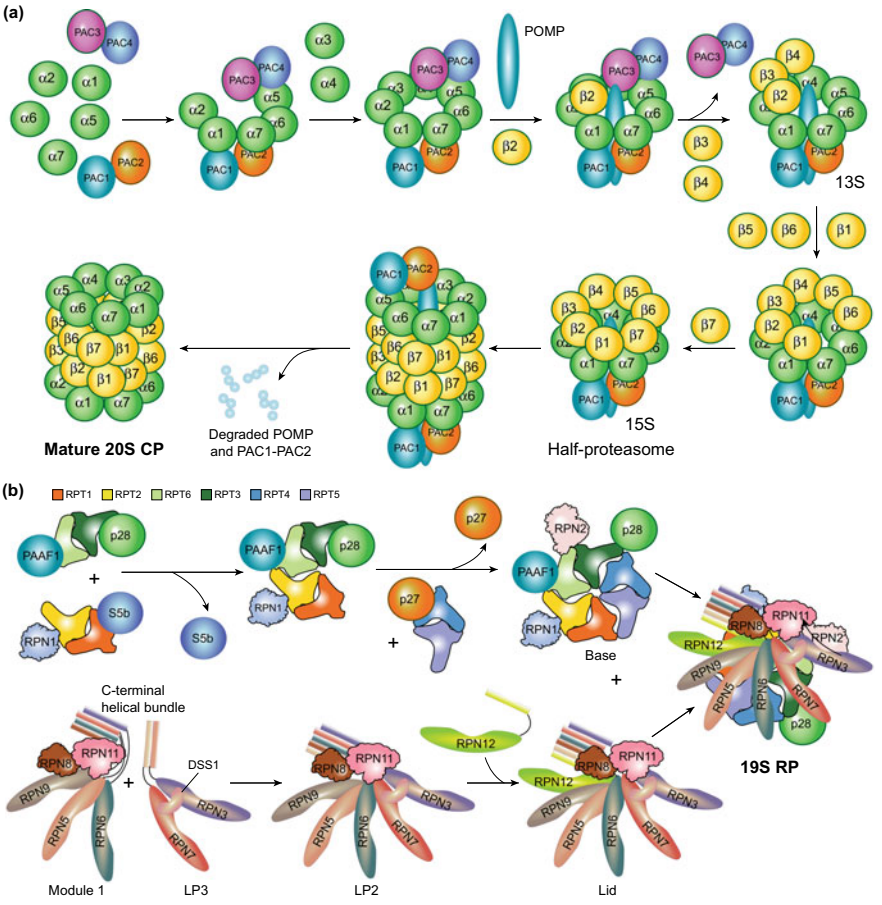


Fig. 1.12 Models of the RP and CP assembly in eukaryotic cells. **a** Assembly pathway of the constitutive CP. The α-ring is assembled with the help of two heterodimers of chaperones: PAC1-PAC2 and PAC3-PAC4. β2 and POMP then associate with the α-ring. The incorporation of β3 and β4 coincides with the dissociation of PAC3-PAC4, followed by sequential incorporation of β5, β6, β1 and β7. Dimerization of two half-CP forms the preholoproteasome. In the last step, the propeptides of the β-ring is auto-cleaved to activate the CP; POMP and PAC1-PAC2 are then degraded to form the mature CP. **b** Hypothetical assembly pathway of the RP. The base assembly is assisted by four RACs: S5b, PAAF1, p27 and p28. The lid can be assembled via a hierarchical pathway without any chaperones. It starts with the formation of two subcomplexes Module 1 and LP3. They associate to form LP2. The final step occurs with RPN12 joining LP2 to form the mature lid. The lid then binds the free base. Note that in yeast the base may first assemble with the CP before its association with the lid

as proteasome biogenesis-associated 1 (Pba1)-Pba4 and underpinning maturation of proteasome 1 (Ump1), respectively.

Four CP assembly chaperones guide the assembly procedure of the α-ring. They form two heterodimers, PAC1-PAC2 (Hirano et al. 2005) and PAC3-PAC4 in human (Pba1-Pba2 and Pba3-Pba4 in yeast) (Hirano et al. 2006; Le Tallec et al. 2007). The PAC1-PAC2 heterodimer promotes the assembly of the α-ring intermediate and prevents premature binding of the RP to the α-ring, as well as aberrant dimerization of α-rings (Hirano et al. 2005; Wani et al. 2015; Stadtmueller et al. 2012; Kusmierczyk and Hochstrasser 2008). Knockdown of PAC1 or PAC2 expression decreases the amount of α-rings and results in accumulation of off-pathway products of aberrant α-ring dimers. PAC1-PAC2 binds to proteasome precursors until the complete formation of the CP. PAC1-PAC2 is degraded by the newly assembled CP and has a half-time of about half an hour ($t_{1/2} \approx 30$ min), which is consistent with the maturation time of the CP.

In yeast, the Pba1-Pba2 heterodimer ensures the appropriate incorporation of the α5 and α6 subunits, whereas the Pba3-Pba4 heterodimer interacts with the α5 subunit of the CP intermediate to ensure the appropriate assembly. Crystal structure of the Pba3-Pba4-α5 complex exhibits pronounced similarity to the structure of the PAC3 homodimer (Yashiroda et al. 2008). Notably, the structures of Pba3 and Pba4 are mostly homologous to the CP subunits. Structural analysis of the Pba3-Pba4-α5 complex suggests that Pba3-Pba4 binds the facet of the α-ring opposite that contacted by Pba1-Pba2, thus sterically clashing with the incoming β4 subunit. This explains why Pba3-Pba4 has to dissociate as β-subunits are incorporated. Deletion of the Pba3 or Pba4 gene causes the accumulation of proteasome intermediates and produces diverse aberrant CP, including those lack of α4 or harboring a second copy of α4 instead of α3 (Velichutina et al. 2004; Kusmierczyk and Hochstrasser 2008; Padmanabhan et al. 2016). Thus, Pba3-Pba4 promotes and guides the proper assembly of the α-subunits, especially α3 and α4.

The α-ring serves as a template for the assembly of the β-ring, which begins with the sequential incorporation of β2, β3 and β4 subunits. POMP/Ump1 is incorporated along with the first β-subunits, whereas the PAC3-PAC4/Pba3-Pba4 dimer is released from the α-ring upon β3 integration (Hirano et al. 2006; Hirano et al. 2008; Li et al. 2007b). The resulting intermediate, known as the 13S complex, then progressively recruits β5, β6, β1 and β7 to form a half CP assembly, also named the 15S complex. Half-CP dimerization is immediately initiated after β7 incorporation (Li et al. 2016b).

The final stage of joining two half CP assemblies together is regulated by POMP/Ump1. POMP prevents premature dimerization of half CP by hindering their dimerization until all β-subunits are incorporated properly (Ramos et al. 1998). After half CP dimerization, the N-terminal propeptides from the β1, β2, β5, β6 and β7 subunits are cleaved in the nascent CP for maturation. Human POMP and PAC1-PAC2 are then degraded by the newly assembled proteasome (Hirano et al. 2006), whereas yeast Pba1-Pba2 is recycled for assembly of newly synthesized CP subunits in yeast (Ramos et al. 1998; Kock et al. 2015). Although there is lack of structural information for POMP/Ump1 interactions, crystal structures of the *Rhodococcus* proteasome mutants with its intact propeptides have offered insights into the final

step of CP maturation (Kwon et al. 2004; Witt et al. 2006). These structures show that the propeptide contacts two adjacent α-subunits, thereby promoting CP assembly and maturation.

RP Assembly Chaperones (RACs)

The lid and base subcomplexes of the RP can assemble separately without and with the RACs, respectively (Fig. 1.12) (Isono et al. 2007). Recombinant lid subunits can self-assemble into a correct subcomplex in the absence of the CP or the base subcomplex in vitro (Lander et al. 2012; Tomko and Hochstrasser 2014). The lid assembly in yeast is preceded by the formation of two intermediate complexes, one named LP3 containing Rpn3, Rpn7 and Sem1, and the other named Module 1 composed of Rpn5, Rpn6, Rpn8, Rpn9 and Rpn11 (Tomko and Hochstrasser 2014; Fukunaga et al. 2010). The intermediate subassemblies LP3 and Module 1 associate to form the subassembly LP2 through the formation of the C-terminal helical bundle, which is finalized by the incorporation of the lid subunit RPN12 (Estrin et al. 2013; Tomko et al. 2015). It remains to be clarified if there are unknown RACs assisting the lid assembly in vivo. The pathway of lid assembly in mammalian cells is not well understood.

Assembly of the base subcomplex is regulated by four RACs in mammalian cells (Table 1.1): S5b (Roelofs et al. 2009; Kaneko et al. 2009), p28 (gankyrin) (Krzywda et al. 2004; Nakamura et al. 2007a), PAAF1 (proteasomal ATPase-associated factor 1) and p27 (Park et al. 2010). Their yeast orthologs are Hsm3, Nas6, Rpn14 and Nas2, respectively. The base assembly is centered around the formation of the heterohexameric ring of AAA-ATPases (RPT1-RPT6). Each RP assembly chaperone binds the C-terminal domain of a distinct RPT subunit in pairs of S5b-RPT1, p28-RPT3 and p27-RPT5, forming three intermediate complexes S5b-RPT1-RPT2-RPN1, p28-RPT3-RPT6-PAAF1 and p27-RPT5-RPT4, respectively (Kaneko et al. 2009; Park et al. 2009; Roelofs et al. 2009; Thompson et al. 2009; Tomko and Hochstrasser 2013; Saeki et al. 2009b). In yeast, an intermediate in base assembly named BP1 has been identified, which contains three Rpts, Rpn1, and Hsm3 (Park et al. 2009). Despite their functional similarity, the four base RACs adopt different structures. Crystal structures reveal that S5b/Hsm3 comprises HEAT repeats (Takagi et al. 2012), that p28/Nas6 comprises ankyrin repeats (Nakamura et al. 2007a), and that PAAF1/Rpn14 forms a WD40 propeller (Kim et al. 2010). The p27/Nas2 sequence predicts a PDZ domain. Modeling of the crystal structure of yeast Hsm3-Rpt1 and Nas6-Rpt3 complexes (Nakamura et al. 2007a, 2007b; Takagi et al. 2012) into the proteasomal ATPase ring suggests that Hsm3 and Nas6 physically occlude the formation of proper RP-CP contacts.

Mechanism of RP-CP Association

There remain two possible assembly pathways for the proteasome holoenzyme. The first pathway is that the CP templates the assembly of the base and the CP-base assembly intermediate further templates the lid assembly for the completion of the 26S assembly. This pathway might dominate in yeast cells, given evidence of observing the base-CP assembly intermediates (Park et al. 2009; Park et al. 2013). The second pathway is that the lid and base assembled into a free RP intermediate, which then assembles with the CP into the holoenzyme. This pathway might dominate in mammalian cells, because the assembly intermediates of the lid with a partial base have been observed in mammalian cells and there is lack of evidence on the base-CP subassemblies in these studies (Lu et al. 2017b; Thompson et al. 2009). Whether the two distinct assembly pathways can coexist and work in parallel in the same system will require further investigations.

Cryo-EM structures of an endogenously purified human p28-RP complex provide important insights into the mechanism of p28/Nas6 in assisting the last step of proteasome assembly and maturation prior to the RP-CP association (Lu et al. 2017b). Surprisingly, the isolated endogenous p28-RP complex samples a very wide conformational landscape. Pronounced conformational dynamics of the complex have been captured by cryo-EM in seven intermediate states designated T_{A1} to T_{A7}, in which the RPT subunits, RPN1 and p28 were reconstructed to subnanometer resolutions whereas the remaining components of RP were solved at 4.6 Å resolution (Lu et al. 2017b). The AAA domains of RPT subunits form a lock-washer-like shape with an opening between RPT2 and RPT6 in all but states T_{A6} and T_{A7}. The opening width is variant in each state, revealing a continuum of conformational changes during an open-to-closed transition of the AAA ring (Fig. 1.12). Remarkably, p28 strongly clashes with the α2 subunit in all conformational states except T_{A7} when the p28-RP is docked into the atomic model of the 26S proteasome. This analysis reveals a functional role of p28 in guiding the CP to select a specific conformational state T_{A7} of the RP to facilitate the last step of proteasome assembly by increasing interfacial complementarity while introducing no significant occlusion. This elegant mechanism is supported by several functional and genetic studies of yeast Nas6 in the RP assembly (Li et al. 2017; Nemec et al. 2019; Park et al. 2013).

A marked structural feature in the free p28-RP lies in the OB ring being blocked by RPN11 that appears to be even tighter than that in states E_D and S_D of the mature proteasome (Lu et al. 2017b). This quaternary arrangement prevents substrates from being prematurely recruited by the AAA unfoldase. Comparison of the p28-RP structure with that of the mature proteasome in the resting state reveals marked conformational changes of the lid upon RP-CP association (Chen et al. 2016a; Lu et al. 2017b). The lid is rotated ~40° around the heptameric axis of the CP. As a consequence, the RPN7 PCI domain is translated ~15 Å towards the CP with repositioning of RPT3 and RPT6, whereas the N-terminal helical repeat region of the RPN6 PCI domain undergoes a prominent rotation of 40° to allow the association of RPN6 with the α2 subunit.

Structural and functional studies have unraveled a conformation-selective, enthalpy-driven mechanism for p28 release upon RP-CP association (Fig. 1.13) (Lu et al. 2017b; Li et al. 2017; Nemec et al. 2019). The presence of p28 facilitates the selection of one p28-RP conformation (T_{A7}) by the CP to form a first-encounter complex of p28-RP-CP, whereas other p28-RP conformations prohibit RP-CP association. The first-encounter complex p28-RP-CP is transient and undergoes pronounced remodeling driven by enthalpic gain at inter-subunit and inter-domain

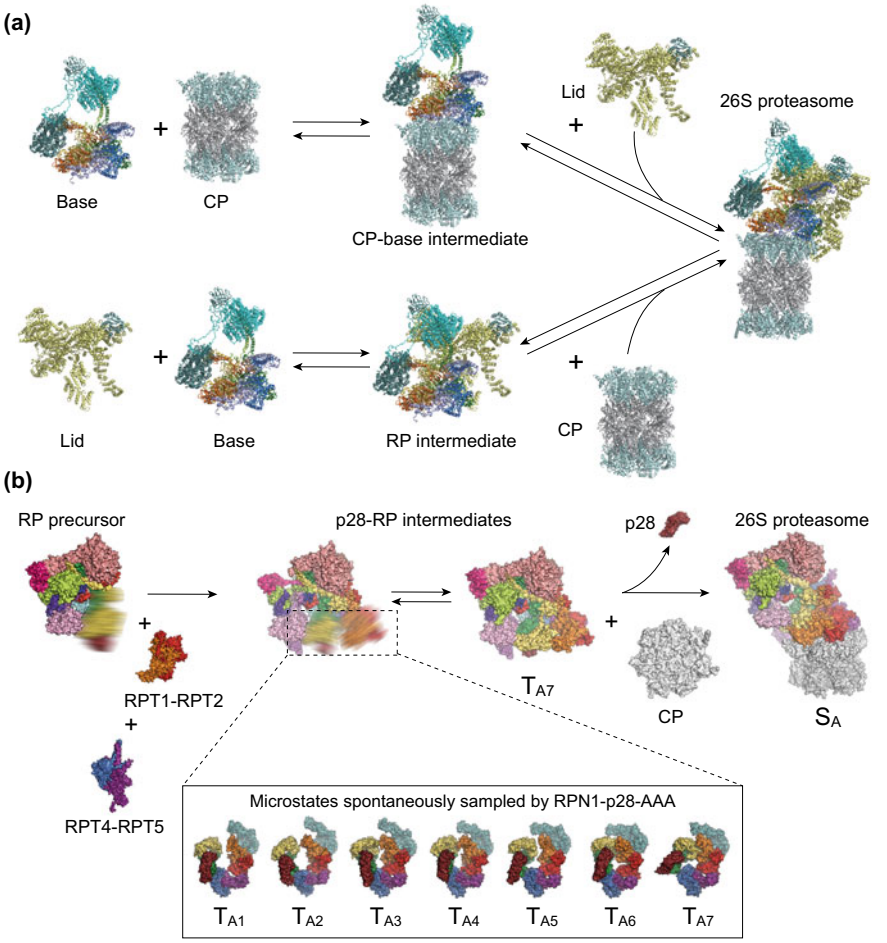


Fig. 1.13 The mechanism of RP-CP assembly. **a** Two possible pathways of RP-CP assembly. **b** A proposed model of the p28 chaperone-guided conformational selection for RP recognition by the CP for proteasome holoenzyme assembly (Lu et al. 2017b). The schematic illustrates the hypothetical CP-independent assembly pathway of the RP prior to the 26S formation. However, this hypothesis is subject to further tests and does not exclude alternative RP assembly pathways (Park et al. 2009; Roelofs et al. 2009)

interfaces. The gap closure between RPT3 and RPT4, which stabilized the ATPase hexamer, and the stronger interaction of the ATPases with the α-ring all contribute to the minimization of the free energy of RP-CP association, which allosterically drives p28 eviction from RPT3 (Lu et al. 2017b). Specifically, RPT3 is translated ~15 Å in order to tighten its association with RPT4, to dock its C-terminal HbYX motif into the α-pocket, and to reposition its p28-binding site toward the α1-α2 surface for p28 ejection. Consistent with this picture, a yeast proteasome mutant *rpt3-Δ1*, in which the C-terminal residue from Rpt3 was deleted, was observed to be unable to release Nas6 (Park et al. 2009), because the mutated Rpt3 loses the affinity of its C-terminal tail with the α-pocket and is unable to complete the Rpt3 remodeling that ejects Nas6. Although there is no structural information regarding other RACs in complex with intermediate base assembly, it is conceivable that this intriguing conformation-selective, enthalpy-driven model might similarly underlie the action mechanism of other RACs.

Cellular Proteasome Regulation

Localization of Proteasome

The cellular proteasome is not only highly dynamic in structure but also in composition and location (Marshall and Vierstra 2019). It could dissociate into free RP and CP subcomplexes and shuttle between the nucleus and cytosol under different physiological conditions during growth, development or proteotoxic stress. Using the fluorescent microscopy, it has been visualized in mammals, plants and yeast that the free RP and CP diffuse throughout both the compartments of nucleus and cytosol (Reits et al. 1997; Enenkel et al. 1998; Russell et al. 1999; Brooks et al. 2000; Pack et al. 2014; Marshall et al. 2015; Gan et al. 2019). In different compartments, the measured proteasome activities exhibit a large variation (Gardner et al. 2005; Chen and Madura 2014; Dang et al. 2016). The local concentration of the 26S proteasome is 830–980 nM in the nucleus of actively dividing yeast cells, but only 140–200 nM in the cytoplasm measured by fluorescence correlation spectroscopy (Pack et al. 2014). Similar results were observed in mammalian neuronal cells (Asano et al. 2015).

Transport of the proteasome between the nucleus and the cytosol could be challenging through the nuclear pores (Beck and Hurt 2017). It was suggested that the nuclear proteasome may dissociate into free CP and RP subcomplexes to enable their exportation separately (Nemec et al. 2017). In proliferating yeast, CP and RP assembly intermediates, each of which carries certain nuclear localization signals (NLS), are imported into the nucleus (Tanaka et al. 1990; Nederlof et al. 1995). An importin-α/β heterodimer comprising Srp1/Kap60 and Kap95, which are two members of the β-karyopherin family, can recognize the NLS carried by a small portion of proteasome subunits and facilitate the transportation (Enenkel et al. 1995). It has been suggested that the CP, lid and base subcomplexes are transported to the nucleus separately by the importin-α/β and the final steps of proteasome assembly

occur in the nucleus (Lehmann et al. 2002; Wendler et al. 2004; Isono et al. 2007). The CP or its assembly intermediates might exist in either import-incompetent or import-competent states, depending on solvent-accessibility of the NLS motifs in certain α-subunits (Tanaka et al. 1990). In support of this hypothesis, cryo-EM studies have shown that NLS sequences are exposed on CP assembly intermediates (Kock et al. 2015; Wani et al. 2015), but are concealed in the mature proteasomes. Import of the base subcomplex into the nucleus was suggested to be mediated by an NLS in RPN2 or RPT2 (Wendler et al. 2004; Isono et al. 2007; Savulescu et al. 2011; Weberruss et al. 2013).

The proteasomes are exported from the nucleus and reversibly sequestered into cytoplasmic proteasome storage granules (PSGs) in quiescent cells (Bingol and Schuman 2006; Laporte et al. 2008; Yedidi et al. 2016; Gu et al. 2017; Marshall and Vierstra 2018b). Supply of a fresh carbon source reverses this process by stimulating rapid import of the free RP and CP back into the nucleus. Blm10 was found to facilitate nuclear import of the mature CP dissociated from PSGs, when cells restart growth from quiescence (Weberruss et al. 2013). In mouse embryonic fibroblasts, the proteasomes of over 3 days old were also observed to be localized in the cytosolic compartment (Tomita et al. 2019). In the green alga *Chlamydomonas reinhardtii*, cryo-ET imaging has found that the proteasomes are enriched at the inner nuclear membrane with a local concentration of over 8 μM around the nuclear pore complexes (Enenkel et al. 1998; Takeda and Yanagida 2005; Albert et al. 2017). Curiously, after neuronal stimulation, a plasma membrane-associated CP was found to specifically degrade ribosome-associated nascent polypeptides associated with the ribosome in a ubiquitin-independent fashion (Ramachandran and Margolis 2017; Ramachandran et al. 2018).

Several studies found that the proteasome can also fulfill nuclear transportation without disassembly (Reits et al. 1997; Chen et al. 2011; Savulescu et al. 2011; Pack et al. 2014), as the size of the nuclear pore complex is large enough to allow free transportation of macromolecular complexes with a diameter of up to 39 nm (Pante and Kann 2002; Burcoglu et al. 2015). To examine this conception, a 26S proteasome is genetically modified by translationally fusing the α4 subunit to RPT1 or RPT2 to stabilize RP-CP association, which did not result in detectable structural defects in the reengineered proteasome. Such a stabilized variant of the 26S proteasome appears to be normally distributed in the nucleus (Laporte et al. 2008; Pack et al. 2014). Because protein synthesis is paused during quiescence, CP intermediates would be absent for nuclear import. This suggests a transport pathway in which the older, stabilized 26S proteasomes are directly imported to the nuclear (Pack et al. 2014).

Post-translational Modifications of Proteasome

Post-translational modifications regulate proteasome assembly, abundance, localization and activity. In total, there are about 350 post-translational modifications detected in the 26S proteasome. These can be categorized into 11 different types of modifications: phosphorylation, *N*-acetylation, *N*-methylation, *N*-myristoylation,

poly(ADP-ribosyl)ation, *O*-glycosylation, oxidation, sumoylation, ubiquitylation, succinylation and proteolytic truncation (Kikuchi et al. 2010; Cui et al. 2014; Hirano et al. 2016; Zong et al. 2014). These modifications provide complex layers of functional regulation that endows the proteasome with exceptional adaptability to many conditions.

Catalyzed by proteasome-associated phosphatases or kinases, phosphorylation is a more extensively studied post-translational modification in the proteasome, and affects many proteasome subunits (Iwafune et al. 2002; Lu et al. 2008; Kikuchi et al. 2010; VerPlank and Goldberg 2017; Liu et al. 2020; VerPlank et al. 2019). The proteasome disassembles into the free CP and RP after treatment with alkaline phosphatase (Satoh et al. 2001). The protein kinase A (PKA) catalyzes phosphorylation of Ser120 in RPT6, which modulates the association of RPT6 and the α2 subunit and can be dephosphorylated by the protein phosphatase 1γ (PP1γ) (Satoh et al. 2001; Asai et al. 2009). Phosphorylation of RPN6 at Ser14 by PKA stimulates multiple proteasomal activities, including peptide hydrolysis by the CP and ATP hydrolysis by the RP, leading to an overall enhanced degradation rate (Lokireddy et al. 2015; VerPlank et al. 2019). The phosphatase UBLCP1 dephosphorylates Ser361 of RPN1 by binding its UBL domain to RPN1, which controls nuclear proteasome assembly by modifying the RP-CP interaction (Guo et al. 2011; Sun et al. 2017; Liu et al. 2020). Similarly, phosphorylation of the α7 subunit regulates the association of Ecm29 with the proteasome (Wani et al. 2016).

Ubiquitylation of the proteasome subunits themselves also provides versatile ways to regulate the proteasome function. For instances, in *Saccharomyces cerevisiae* and *Arabidopsis*, extensive ubiquitylation of the inactive proteasome targets itself for autophagic clearance (Marshall et al. 2015; Cohen-Kaplan et al. 2016). Ubiquitylation of RPT5 seems to regulate an intermediate checkpoint during the base assembly (Fu et al. 2018). Ubiquitylation of the ubiquitin receptors RPN10 and RPN13 downregulates their substrate-binding affinity or orchestrates their association with UBL-UBA receptors (Isasa et al. 2010; Lipinszki et al. 2012; Jacobson et al. 2014; Zuin et al. 2015).

Although numerous studies have started revealing their potential importance in functional regulation of the proteasome, the exact function and mechanism remain unknown for the majority of these post-translational modifications. Besides phosphorylation and ubiquitylation of the proteasome that have been subject to more studies than others as mentioned above, a few additional examples of these efforts are briefly summarized in the following. *N*-myristoylation of RPT2 can tether the proteasome to membrane surfaces (Shibahara et al. 2002; Gomes et al. 2006; Kimura et al. 2012; Kimura et al. 2016). Attachment of *O*-linked *N*-acetylgalactosamine to RPT2 inhibits the activity of the AAA-ATPase motor and reduces the overall proteasome activity (Zhang et al. 2003). In yeast, methylation of the Rpt1 N-terminus is linked to proteotoxic stress induced by the amino acid analog canavanine or hydrogen peroxide (Kimura et al. 2013). *S*-glutathionylation of the α5 subunit appears to regulate the yeast CP gating (Demasi et al. 2003; Silva et al. 2012). The NatB complex

catalyzes *N*-acetylation of the proteasome subunits that regulates proteasome localization in assembling PSGs and cell fitness during aging (van Deventer et al. 2015; Marshall and Vierstra 2018b).

Proteasome-Associated Proteins

Proteasome functions are also regulated by numerous cellular factors that are transiently or reversibly associated with the 26S proteasome (Verma et al. 2000; Wang et al. 2007; Guerrero et al. 2006). These proteins are referred to as proteasome-associated proteins and can be categorized into two groups (Tanaka 2009; Finley 2009). The first group contains proteins related to the ubiquitylation pathway. The extrinsic DUBs USP14 and UCH37 can be recruited to the proteasome by association with RPN1 and RPN13, respectively (Borodovsky et al. 2001; Lee et al. 2016; Verma et al. 2000; Hamazaki et al. 2006; Qiu et al. 2006; Yao et al. 2006). The extrinsic ubiquitin receptors RAD23, DSK2, DDI1 may be considered to belong to this group (Elsasser and Finley 2005; Elsasser et al. 2004; Zhang et al. 2009a). Many E3 ubiquitin ligases, such as UBE3C/Hul5, parkin, and UBE3A/E6AP, are transiently associated with the proteasome (Crosas et al. 2006; Leggett et al. 2002; Sakata et al. 2003; Martinez-Noel et al. 2012; Kuhnle et al. 2018). Other E3s such as Ur1, anaphase-promoting complex/cyclosome (APC/C), SCF^{CDC4}, UFD4 and some E2 enzymes have also been suggested to transiently associated with the proteasome (Demartino and Gillette 2007; Xie and Varshavsky 2000, 2002).

The second group includes other cellular factors that regulate proteasome functions in a non-essential, auxiliary fashion. For example, Ecm29 (Extracellular mutants 29) is a 205-kDa HEAT-repeat-containing protein that recognize a RP-CP intermediate (Leggett et al. 2002). Another example is a proline-rich protein called PI31 (Proteasomal Inhibitor of 31kD) that suppresses proteasome activities by preventing the association of RP and CP and regulates proteasome transport in axons (Zaiss et al. 1999; McCutchen-Maloney et al. 2000; Zaiss et al. 2002; Bader et al. 2011; Li et al. 2014; Minis el al. 2019). In yeast, Ecm29 appears to suppress the ATPase activity of the proteasome (Lee et al. 2011; De La Mota-Peynado et al. 2013). Ecm29 binds the RP-CP intermediate when the CP assembly is prematurely paused due to lack of specific β-subunits (Lehmann et al. 2010). The RP recruits Ecm29 in response to oxidative stress, which induces proteasome disassembly (Park et al. 2011; Wang et al. 2010). It was suggested that the free CP but not the 26S proteasome mediates degradation of oxidized proteins (Davies 2001; Breusing and Grune 2008). Ecm29-regulated disassembly of the proteasome increases the abundance of the free CP and allows cells to adapt to the oxidative stress (Haratake et al. 2016). In addition, by associating with various molecular motors and endosomal components, mammalian ECM29 promotes localization of the proteasome at the ER, the centrosome and likely other cellular locations (Gorbea et al. 2004; Gorbea et al. 2010).

Proteasome Destruction

Autophagic Control of Proteasome

Proteasomes are stable complexes exhibiting a half-life of 16 h in the embryonic fibroblasts of mouse and more than 2 weeks in the liver cells of rat (Tanaka and Ichihara 1989; Pack et al. 2014; Tomita et al. 2019). Their degradation could be also rapid under specific conditions. Three pathways have been discovered to control proteasome turnover (Marshall and Vierstra 2019; Karmon and Ben Aroya 2019). The first pathway is by caspase-catalyzed cleavage. Upon apoptosis induction, RPN2, RPN10 and RPT5 are cleaved by the caspase-3, which impairs proteasome activity in human cells (Sun et al. 2004). In *Drosophila melanogaster*, the caspase-3 cleaves the α2, α4, β4, and RPT1 subunits (Adrain et al. 2004). The second pathway is by the breakdown of dysfunctional proteasome subunits by the UPS itself to avoid their misassembly into the proteasome holoenzyme. It was shown that yeast Hsp42 coalesces the dysfunctional subunits into cytoplasmic condensates that can be removed by the proteasome itself (Peters et al. 2015; Nahar et al. 2019). The third pathway is by proteaphagy that relies on the autophagy system, which degrades large heterogeneous cytoplasmic constituents, such as organelles, protein aggregates and invasive pathogens (Marshall et al. 2015; Marshall et al. 2016; Marshall and Vierstra 2015).

There are two forms of proteaphagy, one that is induced by nutrient starvation and the other that clears inactive proteasomes through ubiquitylation of dysfunctional proteasome itself (Marshall and Vierstra 2015; Marshall et al. 2015; Marshall et al. 2016; Cohen-Kaplan et al. 2016; Waite et al. 2016; Nemec et al. 2017). During the autophagy process, cytoplasmic materials are delivered to the lysosome in mammals or the vacuole in plants and yeast for breakdown by resident hydrolases (Reggiori and Klionsky 2013; Gatica et al. 2018; Marshall and Vierstra 2018a; Levine and Kroemer 2019). During the macroautophagy process, a cup-shaped, isolated membrane named the phagophore is formed de novo and suspended in the cytoplasm. An autophagosome is eventually formed upon the closure of elongated phagophore into a double-layer membrane sphere. The outer membrane of the autophagosome then fuses with the lysosome or vacuole to form an autophagic body. By breaking the inner membrane, the cargos inside the autophagosome are exposed to the entire autophagic body and broken down by lysosomal or vacuolar hydrolases at acidic pH optima, recycling constituents of amino acids, nucleotides, carbohydrates and fatty acids for new biosynthesis (Parzych and Klionsky 2019).

The autophagy is carried out by a series of autophagy-related (ATG) proteins. Upon stimulation by upstream signals from nutrient-responsive kinases, such as Tor1/2 and Snf1, the autophagy is initiated by the serine/threonine kinase complex ATG1. Membrane delivery required by the autophagy is controlled by the transmembrane protein ATG9. Autophagosome nucleation is mediated by the phosphatidylinositol-3-phosphate (PI3P) signal that is generated by the class III phosphatidylinositol-3-kinase (PI3K) complex. Membrane extension is controlled by

the ATG2-ATG18 complex at the site of PI3P labeling. Cargo recruitment is mediated by the ubiquitin-like protein ATG8 (MAP1LC3 or GABARAP in mammals) and its conjugation enzymes in a way highly analogous to ubiquitylation (Ohsumi 2001; Marshall and Vierstra 2018a; Levine and Kroemer 2019). Specifically, ATG8 is activated by the E1 enzyme ATG7, transferred to the E2 enzyme ATG3, and conjugated via an ether linkage to the lipid phosphatidylethanolamine (PE) by a hexameric E3 ligase complex ATG5-ATG12-ATG16. Membrane-bound ATG8 interacts via an ATG8-interacting motif (AIM, also named an LC3-interaction region, LIR) with autophagy adaptors or receptors that recruit the vesicular transport machinery or specific cargo (Rogov et al. 2014; Farre and Subramani 2016; Gatica et al. 2018; Marshall and Vierstra 2018a; Marshall et al. 2019; Noda et al. 2008; Noda et al. 2010; Klionsky and Schulman 2014; Maqbool et al. 2016; Rogov et al. 2017).

Autophagic Clearance of Inactive Proteasome

In plants and yeast, a proteaphagic pathway independent of ATG1 can be stimulated by genetic mutations that impair proteasome assembly, by proteasome inhibitors, such as Bortezomib and MG132, and by pathogenic effectors, such as HopM1 from *Pseudomonas syringae* (Marshall et al. 2015; Marshall et al. 2016; Ustun et al. 2018; Nemec et al. 2017). Proteasome inhibition stimulates ubiquitylation of the proteasome itself and causes its accumulation (Book et al. 2010; Kim et al. 2013; Marshall et al. 2015; Marshall et al. 2016). It is unclear which and how proteasome subunits are ubiquitylated. The ubiquitin receptors in the autophagy system then recognize and deliver the ubiquitylated inactive proteasomes to ATG8. In *Arabidopsis*, RPN10 serves as the autophagy receptor, which uses two distinct UIMs to deliver proteasomes to the autophagosome. One UIM recognizes the ubiquitylated proteasome, while the other UIM recognizes ATG8 in its UIM docking site (UDS) (Marshall et al. 2015; Marshall et al. 2019). The yeast Rpn10 lacks the second UIM for Atg8 binding and has no detectable activity in proteaphagy. Instead, the yeast receptor Cue5 uses a CUE domain to recognize the ubiquitylated proteasome and an AIM domain to recognize Atg8 (Marshall et al. 2016; Lu et al. 2014).

In yeast, aggregation of the 26S proteasome into a superstructure named peri-vacuolar insoluble protein deposit (IPOD) often precedes the ubiquitylation-dependent proteaphagy (Kaganovich et al. 2008; Marshall et al. 2016). The Hsp42 chaperone mediates the formation of IPOD by coalescing aggregated proteins (Specht et al. 2011; Malinovska et al. 2012; Miller et al. 2015). The IPODs are structurally distinct from PSGs (Marshall and Vierstra 2018b; Peters et al. 2016). They can be differentiated based on the co-localized proteins, such as the prion protein Rnq1 in IPODs or Blm10 in PSGs.

The RP-CP association has been observed to be allosterically stabilized through small-molecule inhibitor engagement at the proteolytic active sites in the CP, which appears to induce autophagic degradation of both CP and RP (Kleijnen et al. 2007; Haselbach et al. 2017). By contrast, mutations that compromise the integrity of proteasome assembly induce clearance of the affected CP or RP separately (Marshall

et al. 2016). Thus, the proteaphagy can be carried out for both the proteasome holoenzyme and for the free CP or RP separately. How inhibitor-induced proteaphagy discriminates between functional and dysfunctional proteasomes is very likely linked to the conformational landscape of the proteasome regulated by inhibitor interactions (Haselbach et al. 2017). It is possible that the inhibitor-bound proteasome is strictly stabilized in the conformation of the resting state (S_A) that is better recognized by the ubiquitylation machinery and Hsp42, which directs their accumulation into IPODs, whereas the functional proteasome samples a significantly more expanded energy landscape that can effectively evade proteaphagy machinery (Dong et al. 2019).

Mechanism of Substrate Selection and Ubiquitin Recognition

Degradation Signals

Most proteins targeted for proteasomal degradation are recognized through ubiquitin signals (Komander and Rape 2012), although ubiquitylation is also used for signal transduction in other cellular processes (Kulathu and Komander 2012; Yau and Rape 2016). Another essential requirement for efficient substrate degradation is an unstructured initiation region in the substrate (Prakash et al. 2004; Takeuchi et al. 2007; Zhao et al. 2010). Localization of a protein with such an initiation region to the proteasome could lead to its degradation even in the absence of ubiquitylation (Janse et al. 2004; Erales and Coffino 2014; Murakami et al. 1992). Lack of an effective unfolded initiation site protects a protein from proteasomal degradation (Fishbain et al. 2011; Heinen et al. 2011; Elsasser and Finley 2005). Variations in the length and composition of the unstructured segments affect protein half-life presumably due to the changes of binding affinity with the initial substrate-binding sites in the proteasome (Fishbain et al. 2015; van der Lee et al. 2014). The requirement of an unstructured initiation segment of a substrate has been visualized in the cryo-EM structure of the proteasome in state E_B, whereas the initial ubiquitin recognition steps are observed in the structures of state E_{A1} and E_{A2} (Dong et al. 2019). Although the entire ubiquitin chains have not been visualized in the proteasome, these structural snapshots suggest an intriguing model for consecutive steps in the coordinated ubiquitin recognition and the engagement of the unfolded initiation site with the proteasome, which appears to be also coordinated with the catalytic step of deubiquitylation regulated by the ATP hydrolytic cycles.

Ubiquitin Code

Polyubiquitin chains are the predominant degradation signals in vivo (Fig. 1.14) (Komander and Rape 2012). The compactness and flexibility of the polyubiquitin

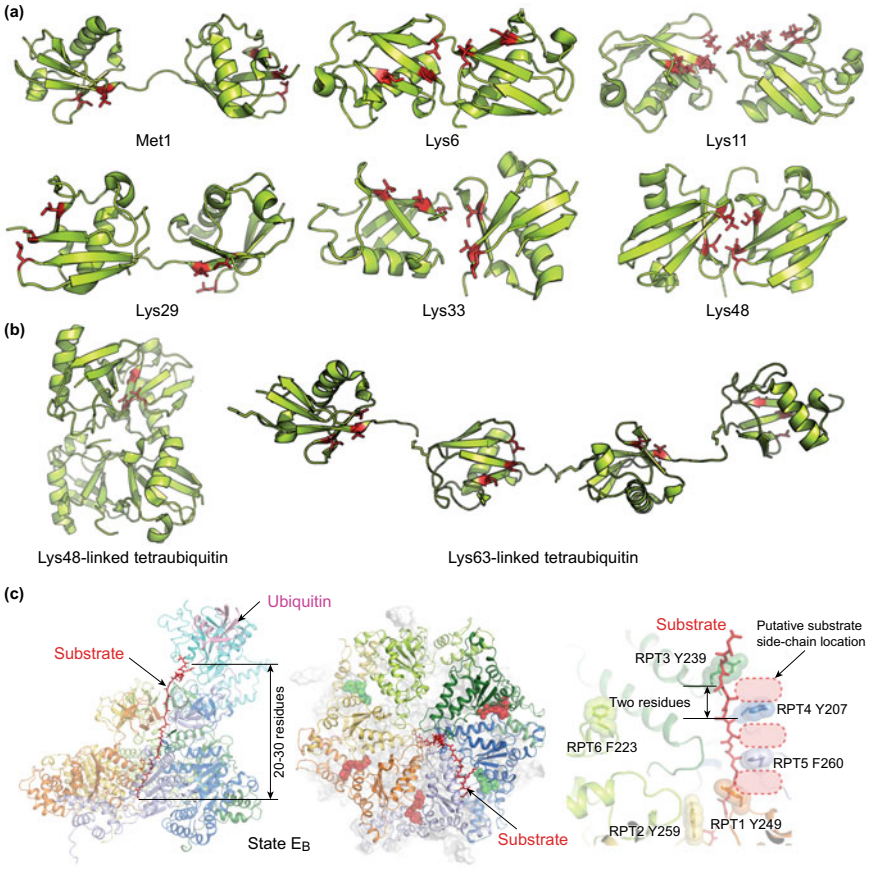


Fig. 1.14 The ubiquitin code and initiation region of substrates. **a** Structures of diubiquitin chains with different linkages. The diubiquitin chains are linked through their N and C-termini (Met1) (PDB ID: 3WXE) (Sato et al. 2015), Lys6 (PDB ID: 3ZLZ) (Hospenthal et al. 2013), Lys11 (PDB ID: 2MBQ) (Castaneda et al. 2013), Lys29 (PDB ID: 4S22) (Kristariyanto et al. 2015a), Lys33 (PDB ID: 4XYZ) (Kristariyanto et al. 2015b), Lys48 (PDB ID: 1AAR) (Cook et al. 1992). **b** Structures of tetraubiquitin chains with two different linkages. The PDB IDs for the Lys48- and Lys63-linked tetraubiquitin chains are 2O6V and 3HM3, respectively (Eddins et al. 2007; Datta et al. 2009). **c** The initiation signal recognized by the AAA-ATPase channels. The ATPase structure in the substrate-bound proteasome in state E_B (PDB ID: 6MSE) is shown from the perspective of side view (left) and top view (middle) (Dong et al. 2019). The right insert shows the close-up view of the substrate interaction with the pore-1 loops, with the aromatic residues in the pore loop highlighted by the transparent sphere representations. The putative location of the substrate sidechains intercalating with these pore-loop aromatic sidechains are marked by dashed red boxes

chains are dependent on the lysine-linkage type and serve as a "ubiquitin code" to control selectivity and specificity of substrate targeting. Ubiquitin-ubiquitin adducts have been observed to form via all possible Lys residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 and Lys63) as well as the N-terminal Met1 in cells by mass spectrometry. The ubiquitin chains can be either formed by homogeneous linkage types or by heterogeneous ones leading to diverse chain topologies including branches. The most commonly detected linkages are through Lys48, followed by Lys11 and Lys63 in yeast (Xu et al. 2009) and Lys63, Lys29, and Lys11 in human (Dammer et al. 2011). The type of chain linkage is determined by the E2 enzyme that works with the E3 ubiquitin ligases or by the class of E3 enzymes that form covalent ubiquitin intermediates (i.e., the HECT and ring-between-ring E3s). Different E2s can work with the same E3 ligase to construct polyubiquitin chains of various linkages and lengths at distinct locations on the same substrate (Ye and Rape 2009). In parallel to their polymerization, ubiquitin chains are trimmed by DUBs associated with the proteasome, or free DUBs in the cytosol (Komander et al. 2009).

The type of chain linkage partly dictates the conformational dynamics of the polyubiquitin signals and thus contributes to the selectivity of substrate targeting by the ubiquitin receptors. The lysine residues in ubiquitin are distributed throughout the molecular surface so that different linkages result in different geometries and flexibility of the polyubiquitin chain. The N- and C-termini are located at opposite ends of the ubiquitin fold, whereas Lys63 is located near the N terminus of ubiquitin (Datta et al. 2009). As a result, both Met1-linked and Lys63-linked polyubiquitin chains adopt extended, highly flexible conformations (Liu et al. 2015; Varadan et al. 2004; Datta et al. 2009). In contrast, Lys48 is positioned roughly halfway between N- and C-termini so that Lys48-linked chains form a more compact, less flexible structure (Eddins et al. 2007). Presumably due to higher conformational entropy of polyubiquitin chains, binding of Met1-linked linear ubiquitin chains with the proteasome is weaker than that of Lys48-linked chains (Thrower et al. 2000). Lys11-linked chains also form a compact structure that is distinct from the structure of Lys48-linked chains (Castaneda et al. 2013).

The ubiquitylated proteins can be recognized by many families of proteins that contain a ubiquitin-binding domain (UBD) (Husnjak and Dikic 2012). The UBD-containing proteins themselves can also be ubiquitylated in certain cases so that ubiquitin regulates the formation of intramolecular interactions to effect specific protein function. At least 20 families of UBDs have been identified (Husnjak and Dikic 2012). Most UBDs have a length between 30 and 120 amino acids and recognize only one or two ubiquitin molecules. The interaction surface on ubiquitin facing its receptors is a hydrophobic patch centered on conserved Ile44, Leu8, and Val70 residues. A smaller area around Phe4, Thr12, and Asn2 mediates the ubiquitin recognition as an endocytosis signal but is not essential in proteasomal degradation (Beal et al. 1996; Sloper-Mould et al. 2001). Due to flexibility of polyubiquitin conformations, chains of different linkages could be recognized by the same UBDs (Alfano et al. 2016). Post-translational modifications of ubiquitin itself have been shown to regulate their interactions with UBDs (Yau and Rape 2016). One notable modification is the phosphorylation of Ser65. Ser65-phosphorylated ubiquitin binds the

proteins at the surface of damaged mitochondria and leads to the clearance of the mitochondria by an autophagic process called mitophagy (Yau and Rape 2016).

Tetraubiquitin chains linked through Lys48 are the canonical signal for proteasomal targeting (Thrower et al. 2000; Pickart 2000; Komander and Rape 2012). The proteasome-binding affinity of Lys48-linked polyubiquitin chain rapidly increases with more ubiquitin units added to the same chain until it becomes a tetraubiquitin chain, beyond which the affinity slowly increases (Thrower et al. 2000). The surfaces of every second ubiquitin molecule at positions i, $i+2$, $i+4$, etc., in Lys48-linked chains are recognized by the proteasome (Thrower et al. 2000). A single tetraubiquitin chain is much more effective in substrate delivery to the proteasome than a single monoubiquitin (Kravtsova-Ivantsiv et al. 2009). However, multiple sites of monoubiquitylation or short polyubiquitylation with different linkages can also lead to efficient degradation both in vitro and in vivo (Hofmann and Pickart 2001; Kirkpatrick et al. 2006; Saeki et al. 2009a; Xu et al. 2009; Kim et al. 2011; Lu et al. 2015; Shabek et al. 2012; Braten et al. 2016; Martinez-Fonts and Matouschek 2016).

Lys63-linked polyubiquitin chains are the second most abundant linkages (14–40% of all ubiquitin chains) in mammalian cells (Dammer et al. 2011; Kaiser et al. 2011) and the third most common chains (16%) in yeast (Xu et al. 2009). In yeast, Lys63-linked chains have been shown to mediate partial degradation of the transcription factor Mga2-p120 (Saeki et al. 2009a). Lys11-linked ubiquitin chains are about 2–5% of all ubiquitin chains in mammalian cells (Dammer et al. 2011; Kaiser et al. 2011), and are as abundant (28%) as Lys48-linked chains in yeast (Xu et al. 2009). They play a crucial role in cell cycle regulation. The E3 ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C) assembles both Lys48- and Lys11-linked ubiquitin chains on many cell-cycle regulators to control their degradation, such as Aurora A, cyclin A, cyclin B1, securin, geminin, Nek2A, Plk1, and E2F1 (Craney and Rape 2013; Peters 2006). On some substrates, APC/C first coordinates with the UBE2C to synthesize short chains with Lys48, Lys63 and Lys11 linkages (Jin et al. 2008). The chains could be further extended by APC/C working with the UBE2S that primarily synthesizes Lys11-linked chains (Williamson et al. 2009; Wu et al. 2010). Interestingly, APC/C-synthesized branched polyubiquitin chains containing both Lys11 and Lys48 linkages are more efficient degradation signals than homogeneous Lys11-linked chains (Meyer and Rape 2014). Lys11-linked chains mediate proteasomal degradation of the hypoxia-inducible factor 1α (HIF-1α) (Bremm et al. 2014), as well as targeted degradation in response to the endoplasmic-reticulum stress (Xu et al. 2009). IRF3 degradation can be targeted by both Lys11- and Lys48-linked chains, which is induced by interferon pathways after viral infection (Qin et al. 2014). In *Drosophila*, the E3 ligase Cul1-Slimb synthesizes similar portions of ubiquitin conjugates with Lys11 and Lys48 linkages on the transcription factor Cubitus interruptus (Ci) targeted for partial degradation by the proteasome (Zhang et al. 2013).

Tandem repeats of the ubiquitin sequence are encoded in the *UBI4* gene in yeast, and in the *UBB* and *UBC* genes in human genome. They express Met1-linked linear polyubiquitin chains that can be cleaved into individual ubiquitin molecules by ubiquitin C-terminal hydrolases. Linear ubiquitin chains can also be synthesized by the

600-kDa linear ubiquitin chain assembly complex (LUBAC) and are used as a scaffold in the assembly of signaling complexes in several pathways (e.g., NF-κB, ERK, Wnt, TNF) (Kirisako et al. 2006; Iwai et al. 2014; Shimizu et al. 2015). The linear polyubiquitin chains do not mediate proteasomal degradation in vitro (Martinez-Fonts and Matouschek 2016). However, a linear chain with a minimum of tetraubiquitin is sufficient for proteasomal targeting assisted by the Cdc48-Npl4-Ufd1 complex in vivo (Kirisako et al. 2006; Zhao and Ulrich 2010), and physiologically mediate the degradation of protein kinase C and TRIM25 (Inn et al. 2011; Nakamura et al. 2006).

The ubiquitin code is not exclusively used by the proteasome but also involved in other proteasome-independent processes such as innate immunity and chromatin remodeling (Komander and Rape 2012; Kulathu and Komander 2012). In yeast, Met4 (Flick et al. 2004) and Cdc34 (Fishbain et al. 2015; Goebl et al. 1994) can be ubiquitylated with Lys48-linked polyubiquitin chains that do not lead to proteasomal degradation. Shorter ubiquitin chains or Lys63-linked polyubiquitin chains are involved in processes like intracellular membrane trafficking during autophagy and endocytosis (Erpapazoglou et al. 2014), DNA repair (Schwertman et al. 2016), cell signaling (Chen and Sun 2009), immune responses (Wu and Karin 2015), mRNA splicing (Song et al. 2010), and the regulation of translation (Silva et al. 2015; Spence et al. 2000). Lys11-linked polyubiquitin chains target MHC-I receptor (Boname et al. 2010) for internalization from the cell surface by endocytosis. Heterotypic linkages mixed with Lys11 and Lys63 on receptor-interacting serine/threonine-protein kinase 1 (RIP1) mediate its binding to the IκB kinase (IKK) complex when the transcription factor NF-κB is activated by TNF-α (Dynek et al. 2010). Monoubiquitylation also functions in membrane trafficking as an endocytosis signal, in the regulation of transcription and in the DNA damage response (Hurley and Stenmark 2011; Ramanathan and Ye 2012).

Initiation Region of Substrates

The presence of an unfolded initiation region is an indispensable prerequisite for initiating substrate degradation by the proteasome (Kraut and Matouschek 2011; Prakash et al. 2009; Yu and Matouschek 2017). Such an unstructured region was found to either flank at the terminus or appear as an internal loop of the substrate. The receptor sites for this initiation region in the proteasome is expected to be the interior of the axial channel of the AAA-ATPase motor in the base (Fig. 1.14c) (Dong et al. 2019). As illustrated in the atomic structure of substrate-bound proteasome in state E_B representing the initial commitment step (Dong et al. 2019), the minimum length requirement for a terminal initiation region spans from the ubiquitin-linking isopeptide bond, across the OB-ring, to the inner end of the ATPase channel, which is equivalent to ∼20–30 amino acids (Fig. 1.14c). This structural estimation is well consistent with several biochemical studies (Prakash et al. 2004; Takeuchi et al. 2007; Verhoef et al. 2009; Yu and Matouschek 2017). The distance of the initiation region from the substrate-conjugated ubiquitin signal apparently determines the

initial affinity of the substrate binding to the AAA channel and hence influences the rate of proteasomal processing (Inobe et al. 2011; Lu et al. 2015; Fishbain et al. 2011). Proteasome-associated proteins such as extrinsic UBL-UBA receptors lacking flexible segments suitable for initial engagement can thus avoid proteasomal degradation. For instance, an unstructured segment intrinsic to RAD23 is not long enough to mediate its recognition by the proteasomal ATPase motor, whereas artificial introduction of a flexible initiation loop at the terminus of RAD23 leads to its proteasomal degradation (Fishbain et al. 2011).

An unstructured initiation loop with a sufficient length is necessary for committing substrates to the proteasome for efficient deubiquitylation and translocation (Peth et al. 2010). This commitment step is crucial for coupling of substrate deubiquitylation with translocation at the proteasome, as shown in the state-E_B structure of the proteasome (Fig. 1.14c) (Dong et al. 2019; Worden et al. 2017). Decoupling of substrate translocation with deubiquitylation can reverse the commitment step, leading to futile substrate processing by the proteasome. The sequence composition of the initiation region has strong correlative effects on the rate of substrate degradation both in vivo and in vitro (Fishbain et al. 2015; Kraut et al. 2012; Yu et al. 2016b). Compared with diverse initiation sequences, those with small sidechains or lower structural complexity (i.e., stretches of glycines, alanines or serines) impair substrate degradation (Hoyt et al. 2006; Sharipo et al. 1998; Yu et al. 2016b). Chemical properties of the initiation region, such as hydrophobicity, electrostatic charge and flexibility further modulate its recognition by the proteasome.

The sequence dependence of the initiation region in the substrate could be explained at least in part by its interaction with the pore loops in the AAA domains of the proteasomal ATPase ring (Fig. 1.14c). In the state-E_B structure of the proteasome, the aromatic residues in the pore-1 loops appear to touch the substrate mainchain backbone and occupy the space between adjacent side chains of the fully stretched substrate polypeptide that assumes a conformation similar to a β-strand. The staircase of the pore-loop aromatic residues intercalating substrate sidechain resembles a "rachet wheel" gripping a "cable tie" via its "teeth". The force generated through this "ratchet teething" model will be dependent of the depth of the "rachet teeth" of the substrate. Large sidechains on the substrate polypeptide chain allow a tighter grip, which can both prevent backsliding and apply a high enough unfolding force. Indeed, introducing a low-complexity sequence of small sidechain residues at the terminus of a highly stable globule domain appears to compromise the gripping force generated by the ATPase motor, thus promoting substrate slippage and releasing partially processed substrate (Hoyt et al. 2006; Sharipo et al. 1998; Kraut and Matouschek 2011; Kraut et al. 2012; Yu et al. 2016b; Tian et al. 2005). Thus, an unstructured but slippery internal loop can serve as degradation-stop signals and allow for partial substrate degradation by the proteasome that then activates transcription factors such as NF-κB, Spt23, and Mga2 (Tian et al. 2005; Piwko and Jentsch 2006; Hoppe et al. 2000, 2001).

Ubiquitin Recognition at the Proteasome

The recruitment of ubiquitylated substrates to the proteasome can be mediated through direct ubiquitin binding with the intrinsic receptors RPN1, RPN10 and RPN13. These intrinsic receptors can also indirectly recruit ubiquitylated substrates by reversibly binding the UBL domains of extrinsic ubiquitin receptors, also known as shuttle factors, that are canonically made of UBL-UBA proteins, such as RAD23, DSK2, and DDI1 families. In yeast, the cellular abundances of the extrinsic and intrinsic receptors are quite comparable (Ghaemmaghami et al. 2003). The intrinsic ubiquitin receptors also provide docking sites for certain DUBs and E3/E4 ligases, which builds up additional layers of complexity in the functional regulation of the proteasome. Binding of either UBL-containing proteins or ubiquitin chains alone, mostly via the intrinsic ubiquitin receptors, are found to generally stimulate the ATPase activity and peptide hydrolysis in the proteasome (Collins and Goldberg 2020; Ding et al. 2019; Kim and Goldberg 2018). This effect suggest that, similar to the influence of substrate engagement, ubiquitin or UBL binding re-equilibrates the ensemble of conformational states of the proteasome toward the E_D-like states (Table 1.2 and Fig. 1.9) (de la Pena et al. 2018; Ding et al. 2019; Dong et al. 2019).

RPN1

The importance of RPN1 is reflected in its ability to recognize both polyubiquitin chains and UBL domains of UBL-UBA proteins. RPN1 is structurally homologous to RPN2, which anchors another ubiquitin receptor RPN13. Both RPN1 and RPN2 resemble a tobacco pipe, with a central proteasome/cyclosome (PC) domain of 11 leucine-rich repeats (LRR) that make up the toroid-shaped "barrel" of the pipe (Fig. 1.15). This toroidal "barrel", composed of only tightly packed α-helices, is demarcated on one side by an N-terminal α-helical domain comprising the wedge-like stem, and on the orthogonal side by a C-terminal domain (CTD) comprising mostly β-stranded barrel-like globular fold (He et al. 2012; Dong et al. 2019). In the toroidal PC domain of RPN1/RPN2, the pore formed by the LRRs is occupied by two α-helices. This packing is mediated by the highly hydrophobic residues in these two central α-helices that sequester them from solvent.

RPN1 and RPN2 are positioned distally from the proteasomal long axis and are the two most surface-exposed RP subunits. RPN1 and RPN2 engage with the RPT1-RPT2 and RPT3-RPT6 CC domains, respectively. In RPN2, both the N-terminal wedge-like domain and the concave side of the PC domain bind the N-terminal CC of RPT3 and RPT6, with the PC domain and CTD being positioned above the entrance of the OB ring. Similarly, the N-terminal CC domains of RPT1 and RPT2 interact with the concave side of the PC domain of RPN1 (Fig. 1.15). The RPN1 PC domain and the RPT1-RPT2 CC domain form a surface cavity that is inserted by a short α-helix from RPN2. This α-helix is the middle part of a long loop (residues 820–871) extending from the RPN2 PC domain. Through such a long-range interaction of RPN1-RPN2,

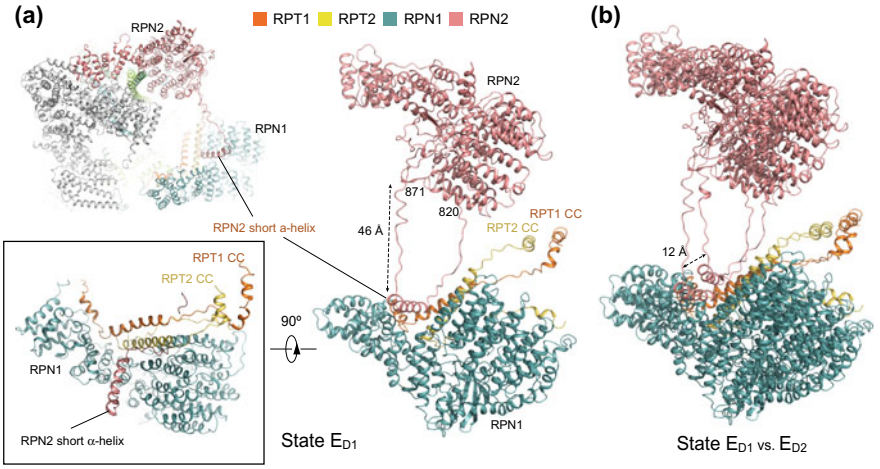


Fig. 1.15 Long-range interaction between RPN1 and RPN2 (PDB IDs: 6MSJ, 6MSK) (Dong et al. 2019). **a** Long-range association between RPN1 and RPN2 via a long looping (residue 820–871) structure extending from the RPN2 PC domain observed in state E_{D1}. **b** Comparison of the RPN1-RPN2 structures between states E_{D1} and E_{D2} shows a prominent translation of RPN1 relative to the CP

a more extensive interface is stabilized between RPT1-RPT2 and RPN1-RPN2 at the concave side of the RPN1 PC domain facing RPN2. Such a quaternary interface is best reconstructed in states E_{D1} and E_{D2}, but less clearly observed in other states ($E_{A\text{-}C}$) (Dong et al. 2019). It appears that RPN1 is most dynamic in state E_C, presumably due to the requirement of ubiquitin release following deubiquitylation.

Structural and mutagenesis studies have revealed that the yeast Rpn1 harbors two receptor sites, designated T1 and T2, which preferentially bind ubiquitin and UBL proteins, respectively (Fig. 1.16) (Shi et al. 2016; Chen et al. 2016b). Although the T2 site was found to only bind UBL proteins such as RAD23, cryo-EM structures of the substrate-bound human proteasome suggest that this site can also engage polyubiquitin (Dong et al. 2019). It is likely that the ubiquitin or UBL-binding affinity is different between the two sites, making one site primary and the other auxiliary under certain scenarios. When the primary site captures a polyubiquitin chain, the auxiliary site could still directly bind a distal ubiquitin molecule in the chain despite lower local affinity.

RPN10

RPN10/S5a seems to represent a "canonical" intrinsic ubiquitin receptor in the proteasome, given its unique proximity to the essential DUB RPN11 than any other intrinsic ubiquitin receptors in the proteasome. RPN10 recognizes polyubiquitin chains via its α-helical UIM domains (Deveraux et al. 1994; Fu et al. 1998; van

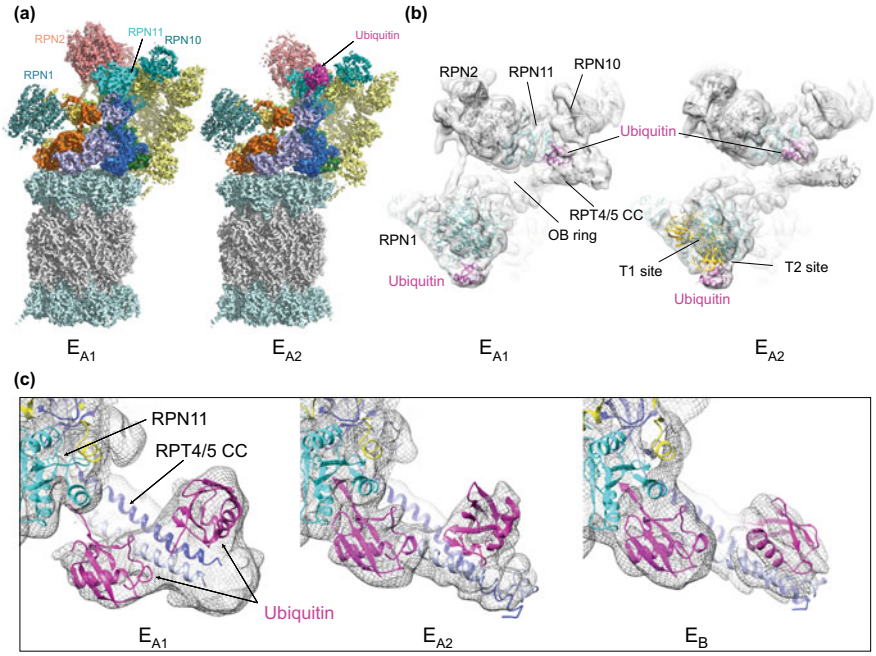


Fig. 1.16 Mechanism of ubiquitin recognition by the 26S proteasome (PDB IDs: 6MSB, 6MSD, 6MSE) (Dong et al. 2019). **a** The cryo-EM maps of states E_{A1} and E_{A2}. **b** Ubiquitin densities in state E_{A1} (left) and E_{A2} (right). The T1 and T2 sites are labelled by fitting the yellow cartoon representation of the NMR structure (PDB ID 2N3U) of RPN1 T1/T2 element in complex with two ubiquitin molecules into our density, showing the ubiquitin on RPN1 is bound to a site very close to the T1/T2 sites (Shi et al. 2016). The density maps are low-pass filtered to 8 Å to show the ubiquitin features clearly, due to the lower local resolution of the ubiquitin density in these maps. **c** Comparison of two ubiquitin moieties between RPN11 and RPT4/5 CC among E_{A1}, E_{A2} and E_B. The cryo-EM densities rendered as grey mesh representations are low-pass filtered to 8 Å. The atomic model of ubiquitin is shown as a magenta cartoon representation

Nocker et al. 1996). RPN10 includes one to three UIMs at its C terminal region. The exact number of UIMs in RPN10 depends on the species. Yeast Rpn10 that possesses only a single UIM still prefers polyubiquitin chains over monoubiquitin (Zhang et al. 2009a, d). Solution NMR structural studies indicate that the two UIMs of the human RPN10 greatly enhance its affinity for a Lys48-linked ubiquitin chain (Zhang et al. 2009d). Although not observed in the context of the proteasome, there is a preference for UIM1 on recognizing the distal ubiquitin and for UIM2 binding the proximal one; when both RPN10 and RPN13 bind the same Lys48-linked polyubiquitin chain, RPN13 seems to recognize the distal ubiquitin whereas the UIMs compete for the proximal ones (Zhang et al. 2009d). This observation is consistent with the spatial arrangement of RPN10 being much closer to the DUB RPN11 and OB ring than RPN13 in the proteasome.

The UIMs are highly dynamic and appear as a poorly resolved diffuse density in cryo-EM structures of the proteasome (Dong et al. 2019). In contrast, the N-terminal VWA domain of RPN10 is tightly folded and well resolved in the cryo-EM structures (Dong et al. 2019; Riedinger et al. 2010). The VWA domain makes extensive contact with the lid subunits RPN8 and RPN9, suggesting that it structurally stabilizes the quaternary architecture of the lid and thus of the RP. Disruption of the RPN10 VWA domain or loss of RPN10 destabilizes the lid-base interaction (Glickman et al. 1998; Fu et al. 2001; Tomko and Hochstrasser 2011). RPN10 also provides primary binding sites to recruit extrinsic ubiquitin receptors like DSK2/Ubiquilin (Matiuhin et al. 2008; Zhang et al. 2009a; Chen et al. 2019) and certain E3 ligases like E6AP/UBE3A (Buel et al. 2020). Binding of the E6AP AZUL (Amino-terminal Zinc-binding domain of Ubiquitin E3a Ligase) domain induces refolding of an unstructured C-terminal segment in RPN10 into a helical bundle (Buel et al. 2020). The yeast Rpn10 is monoubiquitylated in vivo, at Lys71, Lys84 and Lys99 in the VWA domain, which regulates its interactions with substrates by inhibiting the UIM (Isasa et al. 2010). The levels of Rpn10 monoubiquitylation are reversibly controlled by Rsp5, an E3 in the NEDD4 ubiquitin ligase family, and a DUB Ubp2. Interestingly, cold shock, heat shock and cadmium all seem to reduce Rpn10 monoubiquitylation, suggesting that this modification is likely a mechanism of proteasome regulation in response to proteotoxic stress (Isasa et al. 2010).

Although the RPN10 UIMs are poorly resolved in the cryo-EM reconstructions of the human proteasome, there appears to be multiple ubiquitin densities in states E_{A1}, E_{A2} and E_B near RPN10 (Fig. 1.16). The low-resolution density of RPN10 UIMs in state E_{A1} appears to connect with the N-terminal helices of RPT4-RPT5 CC presumably via the polyubiquitin chains bound to the UIMs. Two ubiquitin molecules likely connected within the same chain are attached to the near and far sides of RPT4-RPT5 CC immediately next to the RPN10 UIMs (Dong et al. 2019; Chen et al. 2020). Consistently, in state E_{A2}, one ubiquitin is transferred to the ubiquitin-binding site of RPN11, whereas the density between RPN10 UIMs and RPT4-PRT5 CC appears to be broken (Fig. 1.16b). As the step of substrate deubiquitylation is presumably represented by state E_B, these structural observations suggest that multiple steps of ubiquitin transfer and repositioning or remodeling of substrate-conjugated polyubiquitin chains are required for optimal deubiquitylation (Dong et al. 2019). A flexible link of UIMs to the VWA domain of RPN10 would permit conformational adaptability of the receptor sites in recognizing and delivering polyubiquitin chains of diverse geometries to the deubiquitylation site. In this context, the RPT4-RPT5 CC provides auxiliary receptor sites for enhanced docking of RPN10 UIM-recognized ubiquitin chains and assisting peptide-proximal ubiquitin transfer to the DUB RPN11 for efficient substrate deubiquitylation (Dong et al. 2019; Lam et al. 2002).

RPN13

RPN13 prefers the recognition of the canonical Lys48-linked ubiquitin signals. The N-terminal region of RPN13 comprises the Pleckstrin-like Receptor for Ubiquitin

(PRU) domain that structurally resembles a pleckstrin homology fold (Schreiner et al. 2008; Husnjak et al. 2008). The PRU domain of RPN13 selectively recognizes a dynamic, extended conformation of Lys48-linked diubiquitin via loops extending from the β-strands in the PRU domain and exhibits ~90 nM K_D for Lys48-linked diubiquitin. In the proteasome, the PRU domain binds a flexible C-terminal region of RPN2, which emanates from its LRR-like toroidal domain (Chen et al. 2010; Sakata et al. 2012; Lu et al. 2017a, 2020). The recruitment of RPN13 near the toroid apex of RPN2 positions its PRU domain about four-ubiquitin away from the entrance of the ATPase ring, which is comparable to the distance from the ubiquitin-binding sites in RPN1 and RPN10 (Lu et al. 2020). The PRU domain of human RPN13 is connected by a flexible linker to a C-terminal helical DEUBAD (DEUBiquitinase ADaptor) domain that recruits and activates the DUB UCH37 (Lu et al. 2017a; Sahtoe et al. 2015; Hamazaki et al. 2006; Qiu et al. 2006; Yao et al. 2006). In contrast to the occupancy of the intrinsic receptors typically at more than 50% in the yeast proteasome (Sakata et al. 2012), RPN13 seems to be either substoichiometric or flexibly bound in the endogenously purified human proteasome (Dong et al. 2019).

NMR studies have shown that in the absence of the proteasome, multiple binding modes are feasible for RPN10 and RPN13 in recognition of Lys48-linked tetraubiquitin chains (Zhang et al. 2009d). This prompts the possibility of promiscuous modes of polyubiquitin chain recognition by the intrinsic ubiquitin receptors in the context of the proteasome. It is conceivable that RPN1, RPN10 and RPN13 simultaneously bind polyubiquitin chains to orient substrates for optimal engagement and deubiquitylation (Zhang et al. 2009d; Liu et al. 2019; Lu et al. 2020). This multi-site recognition is both kinetically and enthalpically favored and could be functionally adapted to tackle the vast conformational diversity of the polyubiquitin chains conjugated on protein substrates. Although substrates with multiple short ubiquitin chains can be targeted by any of the known receptors for proteasomal degradation, those delivered to the proteasome via a UBL domain were observed to be degraded more efficiently by RPN13 and RPN1 (Martinez-Fonts et al. 2020). It remains to be clarified if RPN10 and RPN13 can partly function as a substrate shuttle like those extrinsic receptors to some extent.

Delivery of Substrates to the Proteasome

Shuttle Proteins

The extrinsic ubiquitin receptors share common domain organizations. The N-terminal UBL domain that can be recognized by an intrinsic ubiquitin receptor in the proteasome is integrated in a single fold with C-terminal UBA domains that recognize ubiquitin signals (Elsasser et al. 2002; Funakoshi et al. 2002; Schauber et al. 1998; Kaplun et al. 2005; Hofmann and Bucher 1996). Whereas *Saccharomyces cerevisiae* contains one shuttle receptor in each family of Rad23, Dsk2 and Ddi1, higher eukaryotes contain many more paralogs, such as RAD23A/hHR23A,

RAD23B/hHR23B, UBQLN1/2/3/4 and DDI1/2, and other additional UBL-UBA proteins, such as UBAC1, UBL7 and NUB1L. Different paralogs of shuttle-receptor have distinct binding preferences for specific intrinsic ubiquitin receptors owing to differences in their individual UBL interfaces (Chen et al. 2016b). The associations between intrinsic and extrinsic ubiquitin receptors are highly dynamic and complicate their structural analysis. Some UBL-containing proteins can support proteasome delivery in vitro but fail in vivo (Yu et al. 2016a; Gomez et al. 2011; Walters et al. 2002). Differences in the number of UIMs of RPN10 create another level of complexity in the preferential proteasomal binding sites for the UBL domains of the extrinsic receptors. The N-terminal UIM1 of human RPN10 exhibits 25-fold stronger affinity over UIM2 for binding the UBL domain of UBQLN2 (Chen et al. 2019). Recent NMR structural studies on RAD23 and UBQLN2/hPLIC2 examined the structural basis for their preferences to bind the RPN1 and RPN13, respectively (Fig. 1.17) (Chen et al. 2016b; Shi et al. 2016; Chen et al. 2019).

Besides the ubiquitin receptors in the UBL-UBA family, other proteins not in this family have also been found to function as a shuttle factor (Saeki 2017; Finley and Prado 2019). Several members in the ZFAND family of zinc-finger proteins, including ZFAND1, ZFAND2A/B and ZFAND5, are among such candidates (Stanhill et al. 2006; Yun et al. 2008; Osorio et al. 2016; Rahighi et al. 2016; Turakhiya et al. 2018; Hishiya et al. 2006; Lee et al. 2018). Recruiting extrinsic ubiquitin receptors with a flexible domain architecture may allow the proteasome to accommodate a greater variety of ubiquitin signals and substrate structures than those readily recognized by the intrinsic ubiquitin receptors alone. They can also offer substrate selectivity and specificity upstream of ubiquitylation by binding disordered regions in substrates. Using their UBL domains to bind both E4 ubiquitin ligases and intrinsic ubiquitin receptors allows them to ferry ubiquitylated proteins into the proteasome for degradation (Elsasser et al. 2004; Chen and Madura 2002; Kim et al. 2004; Hanzelmann et al. 2010; Tsuchiya et al. 2017). As a result, mistargeted or misfolded proteins that can not be directly recognized by the proteasome due to lack of ubiquitin chains or deficiency of an unstructured initiation region can still be delivered to the proteasome for breakdown (Hanzelmann et al. 2010).

RAD23

RAD23 contains four domains: UBL, UBA1, STI1/XPCB and UBA2 (Walters et al. 2003). The UBA1 and C-terminal UBA2 are separated by the STI1/XPCB (heat-shock chaperone-binding) domain (Yokoi and Hanaoka 2017). UBA1 and UBA2 preferentially bind Lys63-linked and Lys48-linked polyubiquitin chains, respectively, although both can recognize Lys48-linked chains (Raasi et al. 2005). RAD23 UBL domain binds the T1 and T2 sites on the toroidal domain of RPN1 (Elsasser et al. 2002; Shi et al. 2016), positioning it near the entrance of the central channel of the base leading to the CP gate (Rosenzweig et al. 2008). RAD23 may also interact with RPN10 and presumably other ubiquitin receptors via its UBL domain (Ishii et al.

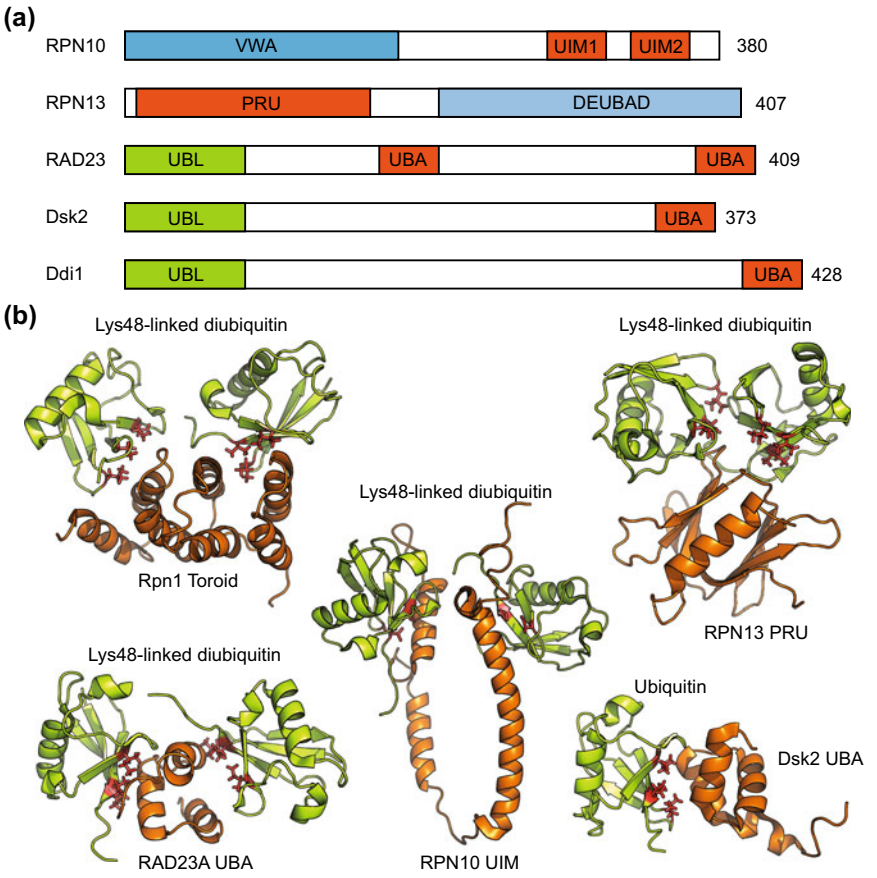


Fig. 1.17 The architecture of ubiquitin receptors and their interactions with ubiquitin. **a** Domain organization of the ubiquitin receptors RPN10, RPN13, RAD23, DSK2 and DDI1. **b** Distinct modes of interaction with ubiquitin are shown for each receptor. Amino acids of each ubiquitin-binding domain that contact ubiquitin are highlighted. The structures correspond to PDB IDs as follows: Ly48-linked diubiquitin-bound yeast RPN1 toroid segment (2N3V) (Shi et al. 2016), Lys48-linked diubiquitin-bound human RPN10 UIM (2KDE) (Zhang et al. 2009d), Lys48-linked diubiquitin-bound human RPN13 PRU (5YMY) (Liu et al. 2019), Lys48-diubiquitin-bound human RAD23A/hHR23A UBA (1ZO6) (Varadan et al. 2005), and ubiquitin-bound yeast Dsk2 UBA (1WR1) (Ohno et al. 2005). PRU, Pleckstrin-like receptor for ubiquitin domain; UBA, ubiquitin-associated domain; UBL, ubiquitin-like domain; UIM, ubiquitin-interacting motif; VWA, von Willebrand factor A domain; UBA, ubiquitin-associated domain

2006). Phosphorylation of its UBL domain inhibits RAD23 interaction with the ubiquitin receptors in the proteasome (Liang et al. 2014). The RAD23 UBL domain can also bind other proteins that can recognize ubiquitin-like structures. For example, it is recognized by the yeast E4 enzyme Ufd2 to facilitate proteasomal degradation of ubiquitin-fusion degradation (UFD) substrates (Kim et al. 2004; Hanzelmann et al.

2010). The peptidyl tRNA hydrolase Pth2 binds the RAD23 UBL, which antagonizes ubiquitin-dependent proteolysis by inhibiting the RAD23 association with the proteasome (Ishii et al. 2006).

The roles of RAD23 in degradation appear to be adaptive and dynamically regulated. Several studies suggest that it suppresses proteasomal degradation under certain conditions (Ortolan et al. 2000; Raasi and Pickart 2003), whereas others have found that it promotes degradation by shuttling protein substrates to the proteasome (Shi et al. 2016; Verma et al. 2004; Elsasser et al. 2004). RAD23 delivers substrates preprocessed by p97/Cdc48 (Olszewski et al. 2019) to the proteasome by protecting the substrates from further processing of their ubiquitin chains, including chain elongation by E4 ligases and deubiquitylation by cytosolic DUBs (Tsuchiya et al. 2017). This is expected to efficiently escort the substrates to the proteasome by preventing premature processing of the ubiquitin signals (Richly et al. 2005). Interestingly, RAD23 is found to be necessary for the formation of nuclear proteasome foci, the transient structures that contain p97/VCP and multiple proteasome-interacting proteins and collectively constitute a proteolytic entre in the nucleus (Yasuda et al. 2020). A liquid-liquid phase separation can be triggered by multivalent interactions of two UBA domains of RAD23 and polyubiquitin chains (Yasuda et al. 2020). RAD23 also functions in the pathway of endoplasmic reticulum-associated degradation (ERAD) by binding of its Rad4-binding STI1/XPCB domain to the deglycosylase Png1, forming a complex that mediates proteasomal degradation of a subset of glycosylated ERAD substrates (Kim et al. 2006). Knockdown of RAD23 genes in mouse models and human cells demonstrated their importance in mammalian development, cell cycle control and apoptosis (Yokoi and Hanaoka 2017).

Ubiquilins/Dsk2

Ubiquilins, the mammalian orthologs of the yeast Dsk2, are a conserved family of four ubiquitin-like proteins (ubiquilin-1-4 or UBQLN1-4) that function as shuttle proteins and deliver ubiquitylated substrates to the proteasome. Similar to RAD23, ubiquilins harbor an N-terminal UBL domain that binds the intrinsic ubiquitin receptors in the proteasome and a C-terminal UBA domain that recruits a substrate (Ko et al. 2004; Hamazaki et al. 2015). Mutations in ubiquilins that compromise their ability to bind the intrinsic ubiquitin receptors were linked to elevated cellular levels of ubiquitylated proteins and aggregate formation, which are implicated in the pathogenesis of several neurodegenerative diseases, such as Huntington's and Alzheimer's diseases, and amyotrophic lateral sclerosis (ALS) (Wang and Monteiro 2007; Wang et al. 2006; Haapasalo et al. 2011; Hjerpe et al. 2016). In yeast, overexpression of Dsk2 impairs proteolysis and exerts a cytotoxic effect (Funakoshi et al. 2002; Biggins et al. 1996; Matiuhin et al. 2008). This effect seems to be attenuated by binding of the Dsk2 UBL to the extraproteasomal Rpn10 UIM that restricts Dsk2 access to the proteasome and alleviates the cellular stress associated with Dsk2 (Matiuhin et al. 2008). This interaction may be regulated by Rpn10 monoubiquitylation (Zuin

et al. 2015), which primarily regulates Rpn10 UIM binding with polyubiquitylated substrates (Isasa et al. 2010; Woelk et al. 2006; Di Fiore et al. 2003).

Ubiquilins have been demonstrated to mediate the degradation of damaged proteins in response to oxidative stress (Liu et al. 2014). Recent studies have revealed a pathway for maintaining protein homeostasis mediated by UBQLN2 (Hjerpe et al. 2016) and a similar role of yeast Dsk2 (Samant et al. 2018) in the breakdown of misfolded or damaged proteins in the nucleus generated by heat shock. The UBQLN2 protein is found to cooperate with HSP70-HSP10 disaggregase chaperone machinery to clear protein aggregates, such as polyglutamine-expanded Huntingtin, via the proteasome (Hjerpe et al. 2016). UBQLN2 recognizes substrate-bound HSP70 and delivers it to the proteasome to enable the degradation of damaged, misfolded or aggregated proteins in the nucleus. The absence of autophagy in the nucleus underscores the importance of this substrate-targeting pathway to the nuclear proteasome. Further studies of similar pathways in yeast suggest that nuclear Dsk2 shuttles Lys48-linked polyubiquitylated substrates to the nuclear proteasome, whereas in the cytoplasm, hybrid Lys11/Lys48 chains enhance the affinity of misfolded proteins by the cytoplasmic proteasome, presumably by engaging multiple ubiquitin receptors at the same time (Samant et al. 2018).

P97/Cdc48

In addition to the extrinsic ubiquitin receptors, other proteins may also participate in shuttling of substrates to the proteasome. The double-ring-shaped hexameric AAA+ ATPase unfoldase p97 or valosin-containing protein (VCP) in mammalian cells, and its yeast ortholog Cdc48, remodel or segregate ubiquitylated substrates for ubiquitin-dependent degradation (Stolz et al. 2011; Xia et al. 2016; van den Boom and Meyer 2018; Meyer et al. 2012). They play an important role in the UPS and protein quality control (Godderz et al. 2015), especially in ERAD (Wolf and Stolz 2012; Christianson and Ye 2014; Wu and Rapoport 2018; Stein et al. 2014) and outer mitochondrial membrane associated degradation (OMMAD) (Taylor and Rutter 2011; Heo et al. 2010; Xu et al. 2011). Although archaeal Cdc48 can assemble with the CP in vitro through artificial crosslinking (Barthelme et al. 2014), similar assembly has not been observed in eukaryotic cells. Since the proteasome requires an unstructured segment in its substrate to initiate substrate processing (Prakash et al. 2004; Fishbain et al. 2011; Inobe et al. 2011), p97/Cdc48 can act upstream of the proteasome to preprocess those substrates that are well-folded without a flexible initiation region or located in membranes (Beskow et al. 2009; Olszewski et al. 2019). p97/Cdc48 can partially unfold the substrate to create an unstructured initiation loop and transfer it to the UBL-UBA proteins like RAD23 and DSK2 that ultimately deliver the substrate to the proteasome for degradation (Itakura et al. 2016; Baek et al. 2011; Richly et al. 2005). Alternatively, p97/Cdc48 might completely unfold the substrate that is directly translocated into the free CP for breakdown (Barthelme and Sauer 2012, 2013). Furthermore, p97/Cdc48 is also involved in other cellular processes, such as ribosomal quality control (Verma et al. 2013; Brandman et al. 2012), extraction

of chromatin-bound proteins (Ramadan et al. 2007; Franz et al. 2016; Dantuma and Hoppe 2012), membrane fusion and vesicular trafficking (Bug and Meyer 2012; Patel et al. 1998), and autophagy (Ju et al. 2009; Papadopoulos et al. 2017).

The p97/Cdc48 complex functions as a molecular motor that mechanically unfolds a substrate translocating across its axial channel. Each p97/Cdc48 protomer comprises two tandem ATPase domains (D1 and D2) that are homologous to the proteasomal RPT subunits, and couple ATP hydrolysis to its pore-1/2 loops that can translocate substrates across the axial channel of the p97/Cdc4 homohexamer (Sauer and Baker 2011; Ogura and Wilkinson 2001; Lupas and Martin 2002; Iyer et al. 2004; Hanson and Whiteheart 2005; Erzberger and Berger 2006; Pamnani et al. 1997). The N domains are not fixed with respect to the double ring (Zhang et al. 2000; Davies et al. 2008). Upon ATP binding to the D1 ATPases, the N domains change from a "down conformation" coplanar with the D1 ring to an "up conformation" above the D1 plane (Banerjee et al. 2016; Tang et al. 2010). To engage a substrate, p97/Cdc48 uses other cofactors such as the Ufd1-Npl4 (UN) heterodimer, which binds the N domains and acts as the ubiquitin receptor (Hanzelmann and Schindelin 2017; Buchberger et al. 2015). To achieve optimal substrate binding, at least five ubiquitin moieties are required in a ubiquitin chain on a substrate (Bodnar et al. 2018; Bodnar and Rapoport 2017). These cofactors endow p97/Cdc48 with the substrate selectivity.

Cryo-EM structures of the yeast Cdc48 in a complex with UN and a ubiquitylated substrate establish the molecular basis for its substrate processing (Cooney et al. 2019; Twomey et al. 2019). Interestingly, a ubiquitin molecule is partially unfolded and the unfolded segment binds a conserved groove of Npl4 leading to the axial channel of the double ATPase rings (Twomey et al. 2019). It is still unclear how the Cdc48-UN unfoldase complex initiates its substrate processing. But the current hypothesis is that the substrate engagement starts on a segment of the ubiquitin molecule proximal to the substrate that can subsequently lead to substrate unfolding. The D1 domains instead of the D2 are suggested to drive the translocation of ubiquitin-substrate adducts after ATP-independent insertion of the ubiquitin N-terminal segment into the axial channel (Bodnar and Rapoport 2017; Olszewski et al. 2019; Twomey et al. 2019). Importantly, this model explains why p97/Cdc48 does not need an unstructured initiation region in the substrate to start on substrate processing.

Proteasome-Associated Ubiquitin Ligases

The proteasome exhibits not only deubiquitylating activity, but in certain condition is also associated with ubiquitin ligase activity. Many ubiquitin ligases have been found to be potentially complexed with the proteasome in a transient fashion, which can impact how a substrate is recognized and processed by the proteasome (Verma et al. 2000; Xie and Varshavsky 2000, 2002; Crosas et al. 2006; Leggett et al. 2002; Besche et al. 2009; Martinez-Noel et al. 2012; Kuhnle et al. 2018). One hypothesis is that there is a mutual regulation between the proteasome and its associated

ubiquitin ligases. The ubiquitylation activity of a ligase could be enhanced once it associates with both a substrate and the proteasome, as the temporal delay between ubiquitylation and deubiquitylation at the proteasome is reduced (Xie and Varshavsky 2000). Ubiquitin-conjugating activity of the yeast Hul5 ligase was proposed to act specifically on proteasome-bound substrates that have already been ubiquitylated by another E3 ligase. Essentially, Hul5 functions as a ubiquitin chain-elongation E4 enzyme regulating the proteasome activity (Crosas et al. 2006). In vitro, Hul5 does not take unmodified cyclin B as a substrate and efficiently adds ubiquitin to cyclin B only if it had already been ubiquitylated (Crosas et al. 2006).

Mechanism of Substrate Deubiquitylation

Once ubiquitin receptors deliver a polyubiquitylated substrate to the proteasome, the ubiquitin signal must be removed from the substrate by the DUBs in the proteasome to enable transport of the unfolded substrate to the proteolytic sites in the CP chamber. Although there are about a hundred distinct DUBs encoded in the human genome, which govern a vast diversity of ubiquitylation pathways (Leznicki and Kulathu 2017), only a few DUBs were found to physically associate with the proteasome. By recycling ubiquitin from substrates, DUBs positively regulate the cellular level of free ubiquitin available for ubiquitylation and, by dynamically reversing ubiquitylation, decide the fate of protein substrates in myriad cellar processes. The isopeptidase activity of proteasome-associated DUBs is often quite low in their isolated form and is prominently enhanced upon their incorporation into the proteasome. The allosteric regulation is reciprocal or bidirectional between the DUBs and the proteasome, since multiple catalytic activities of the proteasome such as the ATPase activity and substrate proteolysis are either stimulated or suppressed in the presence of DUBs. Cryo-EM structure of the proteasome in the act of substrate deubiquitylation and crystal structures of ubiquitin-DUB complexes, along with other cryo-EM studies of DUB in either proteasome subcomplexes or in the absence of substrates, have provided mechanistic insights into DUB activation and regulation by the substrate-proteasome interactions.

Intrinsic Deubiquitinase RPN11

RPN11, a JAMM metalloprotease, is the only intrinsic DUB stably associated with the proteasome and guarding the entrance of the central substrate-translocation channel in the base (Verma et al. 2002). RPN11 is an essential subunit of a functional proteasome and is crucial for cell viability (Maytal-Kivity et al. 2002; Guterman and Glickman 2004). Its role in the proteasome closely resembles the NEDD8 isopeptidase CSN5 in the COP9 signalosome (Cope et al. 2002). RPN11 resides immediately above the OB-ring of the AAA-ATPase motor and is intimately surrounded by the

ubiquitin receptors RPN1 and RPN10. Its ubiquitin-binding site and the auxiliary ubiquitin-binding sites in the RPT4-RPT5 CC domain constitutes a concaved surface resembling a "hook", which is structurally poised for capturing polyubiquitin chains (Dong et al. 2019; Chen et al. 2020). Proximity of RPN11 to the OB-ring and its geometric relationship to the surrounding ubiquitin receptor sites sterically preclude cleavage between ubiquitin moieties within polyubiquitin chains (Dong et al. 2019). Instead, it cleaves ubiquitin chains *en bloc* by hydrolyzing the isopeptide bond linking the C terminus of the first ubiquitin moiety to the substrate lysine (Yao and Cohen 2002).

The MPN domain of DUB RPN11 features three key regulatory regions: Insertion 1 (Ins-1), Insertion 2 (Ins-2) and a catalytic loop. The catalytic Zn^{2+} ion is coordinated with a metal-binding motif ($EX_nHXHX_{10}D$) and is surrounded by the Ins-1 region and the catalytic loop (Verma et al. 2002; Yao and Cohen 2002). The mutations His109Ala and His111Ala at the catalytic site do not impair the proteasome assembly, but inhibit substrate degradation and are cytotoxic in yeast (Verma et al. 2002). The Ins-1 region is located in the vicinity of the RPT4-RPT5 CC domain but does not make direct contact in the resting state of the proteasome. Ins-2 is notably flexible and is not well resolved in the crystal structures of the RPN11-RPN8 MPN-domain dimer (Pathare et al. 2014; Worden et al. 2014). By contrast, Ins-2 assumes an ordered structure and interacts with the toroidal domain of RPN2 in the proteasome (Dong et al. 2019). The catalytic loop of RPN11 contacts RPT4, forming hydrophobic interactions with an OB domain pore loop. These two inter-subunit interactions are critical to stabilize the conformations of Ins-2 and the catalytic loop in the pre-engaged state of the proteasome.

The lid subunit RPN8 and RPN11 form a heterodimer in the proteasome and when purified in isolation. In the free RPN11-RPN8 dimer, the Ins-1 region of RPN11 adopts a loop conformation that blocks the catalytic active site and self-inhibits the DUB activity (Worden et al. 2014; Pathare et al. 2014). In the free lid subcomplex, RPN5 further stabilizes the Ins-1 loop in the inhibited state, preventing RPN11 activation until it is assembled into the proteasome (Dambacher et al. 2016). In contrast, several related DUBs, such as Sst2 and AMSH-LP, remain active in isolation and their Ins-1 regions adopt a β-hairpin conformation, making the active site accessible (Sato et al. 2008; Davies et al. 2011; Shrestha et al. 2014). RPN11 retains a basal level of DUB activity in the context of the free RP subassembly (Lu et al. 2017b). In the free RP, RPN11 blocks the substrate entrance of the OB ring similar to its conformation in state E_D of the proteasome, which prevents the OB ring from accepting substrates. The free RP exhibits a moderate ATP-independent DUB activity and its lack of any substrate-unfolding activity (Lu et al. 2017b). The 'closed' OB-ring in the free RP minimizes this risk of premature substrate deubiquitylation and unfolding prior to incorporation of the free RP into the mature proteasome. It remains to be clarified what cellular roles the basal DUB activity of RPN11 will play in the free RP.

Mechanism of Substrate Deubiquitylation by RPN11

The atomic structure of the substrate-bound proteasome in state E_B reveals how the DUB activity of RPN11 is dependent on ATP hydrolysis and is coupled to substrate unfolding (Fig. 1.18) (Dong et al. 2019). In state E_B, the ubiquitin-bound RPN11 interacts with both RPN8 and RPT5, which forms a subcomplex. This quaternary interface already starts to form in state E_{A2}, when the substrate is not yet captured by the ATPase ring. Comparison of ubiquitin densities between states E_{A1}, E_{A2} and E_B suggests that the ubiquitin directly linked to the substrate is transferred from RPT4-RPT5 CC to RPN11 (Dong et al. 2019). In both states E_{A2} and E_B, RPN11 does not interact with the substrate on the peptide-proximal side of the scissile isopeptide bond, explaining its non-specificity of removing ubiquitin chains from any ubiquitylated substrates.

The ubiquitin-binding interface is located at a hydrophobic pocket around Trp111 and Phe133 of RPN11 (Worden et al. 2014; Pathare et al. 2014; Dong et al. 2019).

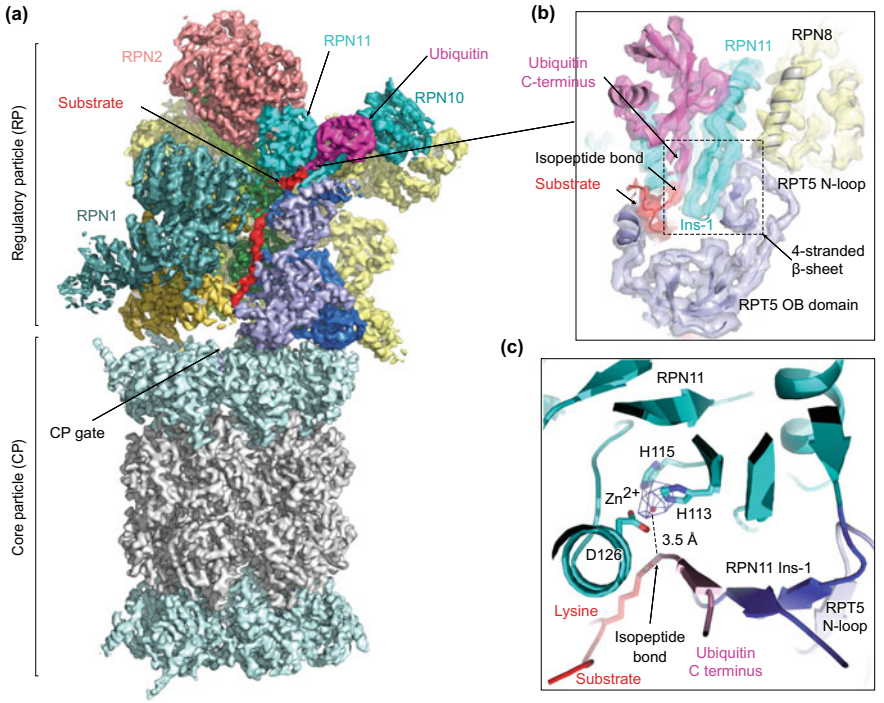


Fig. 1.18 Mechanism of substrate deubiquitylation by the 26S proteasome (PDB ID: 6MSE) (Dong et al. 2019). **a** Cryo-EM density maps of substrate-bound human proteasome in state E_B at 3.3 Å. **b** A close-up view of the quaternary interface around the scissile isopeptide bond between ubiquitin and the substrate lysine in E_B. The cryo-EM density is rendered as a transparent surface, superimposed with the cartoon representation of the atomic model. **c** A close-up view of the zinc ion (hotpink sphere) closely approached by the isopeptide bond. The zinc ion density is shown as a blue mesh at 10σ level. Side chains of Rpn11 coordinating with the zinc ion are labeled

The Ins-1 loop adopts a β-hairpin conformation, and pairs from one side with the C-terminal strand of ubiquitin and from the other side with a segment of the RPT5 N-loop (residues 99–119) emanating from the OB ring. Together they form a four-stranded β-sheet. This β-sheet places the isopeptide bond right next to the zinc ion in the RPN11 active site (Dong et al. 2019). The Ins-1 β-hairpin is rotated outward in state E_B relative to its conformation in the crystal structure of ubiquitin-bound RPN11 and is better aligned with the zinc-binding site in the proteasome (Worden et al. 2017). The conserved RPT5 N-loop stabilizes the ubiquitin contact with RPN11 and adjusts the orientation of the isopeptide bond for efficient deubiquitylation, whereas it is unstructured in most other states (E_{A1}, E_{C1}, E_{C2}, E_{D1} and E_{D2}).

The Ins-1 region of RPN11 keeps changing its conformation with different states of the proteasome (Fig. 1.19). The Ins-1 remains as a β-hairpin conformation throughout states E_{A2}, E_B and E_{C1}, whenever ubiquitin is bound. It becomes a large open loop in state E_{A1} and refolds into a closed, tighter loop in states E_{C2}, E_{D1}, and E_{D2} whenever ubiquitin is released. The quaternary organization surrounding the zinc-binding site of RPN11 appear to explain why the DUB activity of RPN11 is prominently upregulated in the proteasome than in its non-proteasome forms, such as the free RP assembly (Lu et al. 2017b), the free lid subcomplex (Dambacher et al. 2016) and the heterodimeric RPN8-RPN11 form (Pathare et al. 2014; Worden et al. 2014). The dynamic regulation of Ins-1 conformations allows allosteric coordination and timing between substrate commitment, deubiquitylation and degradation. Initial substrate engagement with the AAA-ATPase motor allosterically regulates the conformational changes of RPN11 and RPT5 N-loop and enables deubiquitylation only for committed substrates. Indeed, disrupting the closed state of Ins-1 stimulates the RPN11 DUB activity but also causes degradation failure, because ubiquitin chains are prematurely removed, allowing substrates to escape from the proteasome prior to translocation and degradation (Worden et al. 2017).

The structural revelation of state E_B fills up a fundamentally critical intermediate, allowing the observation of stepwise activation of the CP by docking of the RPT C-termini into the α-pockets one at a time (Dong et al. 2019). During the process of CP gate opening, the relative positions of the RPN11 catalytic site, the central pore of the ATPase ring, and the CP gate are coaxially aligned (Matyskiela et al. 2013; Sledz et al. 2013; de la Pena et al. 2018; Dong et al. 2019), preparing the proteasome for processive substrate translocation into the CP chamber.

Deubiquitinase USP14

USP14 or its yeast ortholog Ubp6 is a cysteine-dependent DUB that is reversibly associated with the proteasome (Borodovsky et al. 2001). The N-terminal UBL domain of USP14 binds RPN1, whereas its catalytic USP domain contacts the exterior of the OB and AAA rings opposite RPN11 (Bashore et al. 2015; Shi et al. 2016; Huang et al. 2016; Aufderheide et al. 2015; Leggett et al. 2002). Such multivalent interactions facilitate stimulation of its DUB activity by lifting off the inhibitory loops blocking

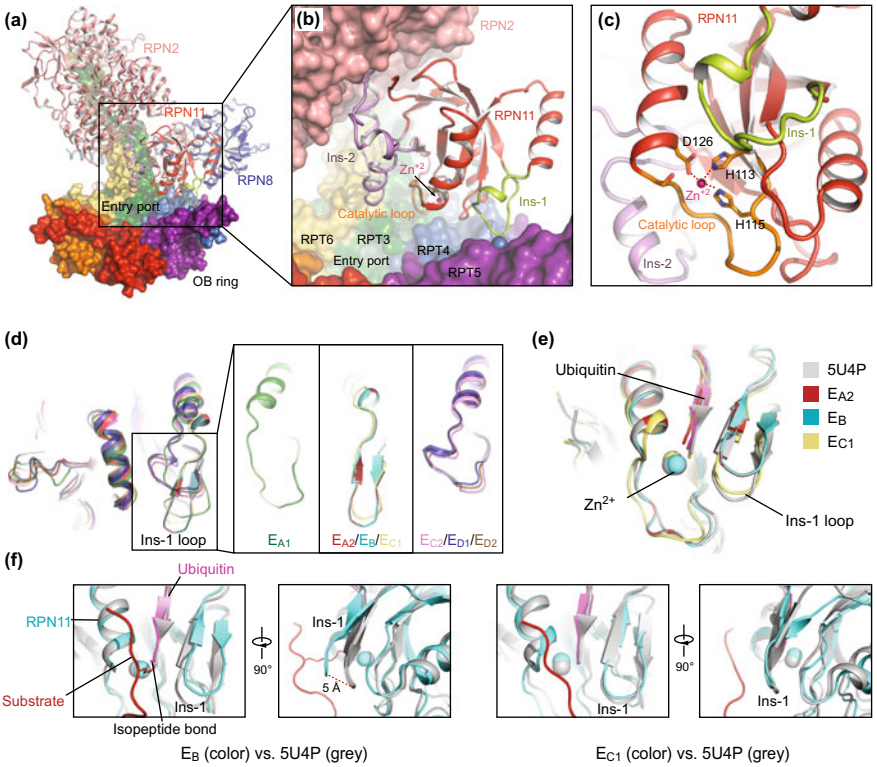


Fig. 1.19 Conformational dynamics of the RPN11 associated with an enhanced isopeptidase activity in the proteasome (PDB IDs: 6MSB, 6MSD, 6MSE, 6MSG, 6MSH, 6MSJ, 6MSK) (Dong et al. 2019). **a** Overview showing the inter-subunit relationship between RPN2-RPN11-RPN8 and the CC-OB domains of the ATPases in state S_A. Black box indicates the location to which the close-up view in panel (**b**) is zoomed. **b** Close-up view of Rpn11 in cartoon representation. The adjacent RPT subunits and RPN2 are shown in surface representation. **c** Close-up view of the Zn^{2+} active site of Rpn11. The viewing angle is rotated clockwise ~90° relative to that of panel (E). **d** Comparison of the Ins-1 loop of RPN11 in different states. **e** Comparison of the RPN11 structures in states E_{A2}, E_B and E_{C1} around the zinc-binding site and Ins1 region with that in the crystal structure (PDB ID: 5U4P) of ubiquitin-bound RPN11-RPN8 complex from yeast (Worden et al. 2017). **f** Closeup comparison of the RPN11 Ins1 structure between state E_B and 5U4P (left two panels) and between state E_{C1} and 5U4P in two orthogonal perspectives, showing a 5-Å displacement of the Ins-1 β-hairpin in E_B relative to 5U4P or E_{C1}. This displacement is not observed between E_{C1} and 5U4P, suggesting that the Ins-1 β-hairpin tilt in E_B is mostly to optimize the coordination of isopeptide bond with the zinc ion

the active site. Thus, the DUB activity of USP14 is allosterically regulated by its interaction with the proteasome and is tightly coupled with the conformational state of the proteasome (Bashore et al. 2015; Hanna et al. 2006). Association with the proteasome activates USP14/Ubp6 ~300- to 800-fold when assayed with the model substrate ubiquitin-AMC (Leggett et al. 2002; Lee et al. 2010). Re-equilibrating the yeast proteasome conformation toward the open-gate states with ATPγS further

increases the activity of Ubp6 by a couple of folds in cleaving the ubiquitin-AMC substrate (Bashore et al. 2015). The active sites of USP14 and RPN11 are only 35 Å apart in the base, allowing USP14 to antagonize ubiquitin binding to RPN11 (Bashore et al. 2015; Hanna et al. 2006). Ubiquitin-bound USP14 appears to stimulate the proteasomal ATPase activity and CP gate opening, suppress substrate degradation by destabilizing the resting state of the proteasome, inhibit deubiquitylation by RPN11 and the basal peptide hydrolysis, and surprisingly enhance the E4 ligase UBE3C/Hul5 association with the proteasome (Bashore et al. 2015; Aufderheide et al. 2015; Peth et al. 2009; Leggett et al. 2002; Kuo and Goldberg 2017; Kim and Goldberg 2017). Curiously, the UBL domain of USP14 alone can stimulate the proteasome activation (Kim and Goldberg 2018). Hence, USP14/Ubp6 may allosterically regulate the proteasome activities at multiple levels, including the recognition of ubiquitin or substrate, timing of deubiquitylation and processivity of substrate proteolysis.

When the USP domain is recruited to the base of the RP, the OB-ring is adjacent to the proximal ubiquitin binding site of USP14 (Huang et al. 2016). This resembles the quaternary structural relationship between the OB-ring and, in its immediate vicinity, the proximal ubiquitin-binding site on RPN11 (Dong et al. 2019). Thus, the biochemical behaviors of RPN11 and USP14 with respect to an incoming substrate might be comparable and somewhat competitive to each other. With a preference for cleavage at the substrate-linked isopeptide of ubiquitin chains, they both could reverse ubiquitin modifications *en bloc* on a time scale of milliseconds to seconds (Lee et al. 2016). Interestingly, USP14 can rapidly cleave ubiquitin from a substrate that is conjugated to multiple polyubiquitin chains, whereas deubiquitylation of single-chain conjugates or disassembly of free polyubiquitin chains by USP14 is very slow (Lee et al. 2016). Curiously, USP14 can cleave long, unbound Lys48-linked chains better than does RPN11 for certain substrates (Mansour et al. 2015). Thus, USP14 seems to be specialized in removing supernumerary ubiquitin chains and cannot replace RPN11 in thorough deubiquitylation of substrates (Lee et al. 2016). In contrast to the essential role of RPN11 in proteasomal activities, inhibition or absence of USP14/Ubp6 in the proteasome accelerates substrate degradation in vitro (Hanna et al. 2006) and is tolerable for yeast cells (Guterman and Glickman 2004). However, the absence of Ubp6 causes a growth deficiency due to enhanced proteasomal degradation and depletion of free ubiquitin (Leggett et al. 2002). USP14 deficiency in mouse embryonic fibroblasts increased the abundance of UCH37 and RPN13 in the proteasome, suggesting that UCH37 may compensate the loss of USP14 (Kim and Goldberg 2017).

Deubiquitinase UCH37

UCH37 is another cysteine-dependent DUB reversibly associated with the proteasome in higher eukaryotes, such as *Homo sapiens* and *Drosophila melanogaster*. Its ortholog Uch2 was found in *Schizosaccharomyces pombe* but not *Saccharomyces*

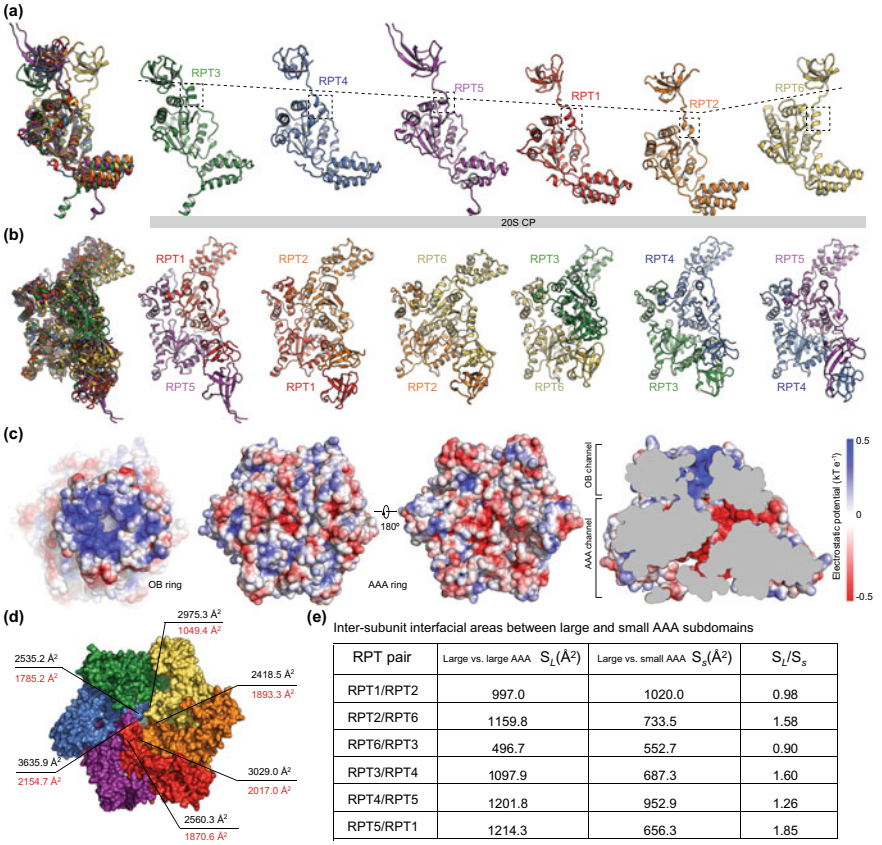
cerevisiae (Lam et al. 1997; Holzl et al. 2000). UCH37 consists of two domains, an N-terminal catalytic domain representative of the UCH family and a fibrous C-terminal tail (Burgie et al. 2012). UCH37 is activated by binding of its C-terminal tail to the N-terminal DEUBAD domain of RPN13 in the proteasome but not in a free UCH37/RPN13 complex (Yao et al. 2006; Sahtoe et al. 2015; Vander Linden et al. 2015). A loop sequesters the catalytic cysteine (Cys88) in the free UCH37 (Burgie et al. 2012). Upon association with RPN13, this loop is remodeled to open the access to the catalytic triad residues (Cys88, His164 and Asp179), which gives rise to an approximately five-fold increase in its affinity for ubiquitin (Yao et al. 2006; Hamazaki et al. 2006; Qiu et al. 2006; Sahtoe et al. 2015; Vander Linden et al. 2015). Within the free RPN13, the DEUBAD domain interacts with the PRU domain and sterically occludes its ubiquitin-binding site. Association of RPN13 with RPN2 in the proteasome modifies this inter-domain interaction and makes the PRU domain to become accessible for ubiquitin binding, leading to activation of UCH37 (Chen et al. 2010; Lu et al. 2017a).

Once activated in the proteasome, UCH37 can catalyze the removal of distal Lys48-, Lys11-, and Lys6-linked ubiquitin chains conjugated to substrates (Lam et al. 1997). Inadequately ubiquitylated proteins could be released by UCH37 from the proteasome. UCH37 may also reverse regulatory ubiquitylation modifications from certain proteasome subunits (Jacobson et al. 2014). It might trim distal ubiquitin moieties, release them from the ubiquitin receptors and vacate the ubiquitin-binding sites in the proteasome (Zhang et al. 2011). Despite its putative location that is about the length of tetraubiquitin away from the OB ring of the AAA-ATPase motor (Holzl et al. 2000; Dong et al. 2019; Lu et al. 2020), UCH37 association with the proteasome can allosterically stimulate the proteasomal ATPase activity and CP gate opening (Peth et al. 2013a). These allosteric effects are in line with those seen in the USP14-bound proteasome and further underscore the global allosteric coupling between all components of the proteasome. In addition to its role in the proteasome, UCH37 also appears as a subunit of the chromatin-remodeling complex INO80 (Yao et al. 2008). It remains to be clarified what substrate specificities of UCH37 are and how UCH37 acts on its substrates in the proteasome. How the three DUBs, RPN11, USP14 and UCH37, are coordinated and collectively regulated in the context of the proteasome will be an intriguing open question awaiting answers.

Operating Principles of Proteasomal AAA-ATPase Motor

Proteasomal AAA-ATPase Motor

The proteasome utilizes its ATPase motor module to mechanically unfold the globular domains of substrates and deliver them into the CP for their breakdown into short polypeptides. The proteasomal ATPases are members of the classic AAA clade in the AAA+ superfamily (Fig. 1.20) (Vale 2000; Smith et al. 2006). Other members in



(e) Inter-subunit interfacial areas between large and small AAA subdomains

RPT pair	Large vs. large AAA S_L(Å²)	Large vs. small AAA S_s(Å²)	S_L/S_s
RPT1/RPT2	997.0	1020.0	0.98
RPT2/RPT6	1159.8	733.5	1.58
RPT6/RPT3	496.7	552.7	0.90
RPT3/RPT4	1097.9	687.3	1.60
RPT4/RPT5	1201.8	952.9	1.26
RPT5/RPT1	1214.3	656.3	1.85

Fig. 1.20 Asymmetric structure of the ATPase heterohexamer in the S_A/E_A state (PDB ID: 6MSB) (Chen et al. 2016a; Dong et al. 2019). **a** The OB-AAA domain structures of RPT subunits in a cartoon representation from a perspective showing an L-shape. The leftmost panel shows the six RPT structures superimposed, with each structure separately shown on the right. **b** The dimeric AAA domain structure of two adjacent RPT subunits. The leftmost panel shows the six AAA dimer structures superimposed together, with each structure separately shown on the right. **c** The electrostatic structure of the OB ring (left), AAA ring (middle), and ATPase channel (right). **d** The inter-subunit interfacial areas between two adjacent RPT subunits. The black numbers show the total inter-subunit interfacial areas, whereas the red numbers show the interfacial areas between the adjacent AAA domains. **e** A table shows the inter-subunit interfacial areas between the large and small AAA subdomains from two adjacent RPT subunits. Note that the two adjacent small AAA subdomains do not contact each other

this clade include FtsH family, Cdc48 family, and ClpA/B/C-Domain 1 (D1) family (Ogura and Wilkinson 2001; Lupas and Martin 2002; Iyer et al. 2004; Erzberger and Berger 2006; Hanson and Whiteheart 2005). The HCLR clade, which includes the HslU/ClpX family, ClpA/B/C-Domain 2 (D2) family, and Lon family, and features a pre-sensor 1 insertion residing in at least one AAA+ module, is found in many

protease complexes and functions as substrate unfolding machinery resembling the proteasomal AAA-ATPase motor (Sauer and Baker 2011).

The AAA domain of RPT contains highly conserved motifs commonly observed in the AAA+ superfamily proteins, including Walker A, Walker B, inter-subunit signaling (ISS) motif, sensor 1, arginine finger (R-finger), sensor 2, and pore-1/2 loops (Fig. 1.21) (Sauer and Baker 2011; Ogura and Wilkinson 2001; Lupas and Martin 2002; Iyer et al. 2004; Hanson and Whiteheart 2005; Erzberger and Berger 2006; Chang et al. 2017; Zhang and Mao 2020; Puchades et al. 2020). The nucleotide-binding pocket is intimately surrounded in "cis" by Walker A, Walker B, sensor 1, and sensor 2 within the same AAA domain, and two R-fingers and ISS motif in "trans" from the large AAA subdomain of the clockwise adjacent ATPase subunit (Wendler et al. 2012). With these structural motifs, ATP hydrolysis is coupled to conformational changes of the ATPases that apply mechanical work on a substrate through the axial channel-aligning pore loops.

Conformational Dynamics of the AAA-ATPase Motor

The Substrate Translocation Pathway

The substrate-translocation pathway in the proteasomal ATPase motor is narrowed by the inward-facing pore loops that are common to the classic clade ATPases (Fig. 1.22) (Erales et al. 2012; Beckwith et al. 2013; Hinnerwisch et al. 2005; Martin et al. 2008; Dong et al. 2019). The AAA channel exhibits a right-handed helical architecture with a much narrower constriction compared to the OB channel in the resting state (S_A) of the proteasome. The interior of the AAA channel is largely negatively charged, in contrast to the positively charged interior of OB channel (Chen et al. 2016a; Dong et al. 2019). This may reflect distinct roles of the OB and AAA domains in priming and mechanically translocating substrates. Despite this difference, the OB and AAA channels share a dramatic enrichment of tyrosine residues. The OB channel features six tyrosine residues (Tyr147 in RPT1, Tyr72 and Tyr121 in RPT6, Tyr111 in RPT3, Tyr79 in RPT4, and Tyr158 in RPT5), whose sidechain oxygen atoms point towards the substrate-translocation pathway. Likewise, the AAA channel is also decorated with five tyrosine residues in the resting state. These tyrosine residues may facilitate substrate engagement and translocation via cation-π or π-π interactions (Glynn et al. 2009).

The pore-1 loops of the RPT subunits feature the conserved '[Tyr/Phe]-[Val/Leu/Ile]-Gly' sequence pattern and arrange into a super-helical staircase. In many ATP-dependent AAA unfoldases, such as ClpX, HslU, LonA, FtsH and PAN (Glynn et al. 2009; Iosefson et al. 2015), their homologous pore-1 loops drive substrate translocation. The pore-2 loops constitute a second super-helical staircase running in the opposite of the pore-1 loop staircase (Fig. 1.22), which may function a 'sensor' for the state of the substrate translocation as these pore-2 loops are directly extended from the Walker B motif and can translate the presence of substrate in the

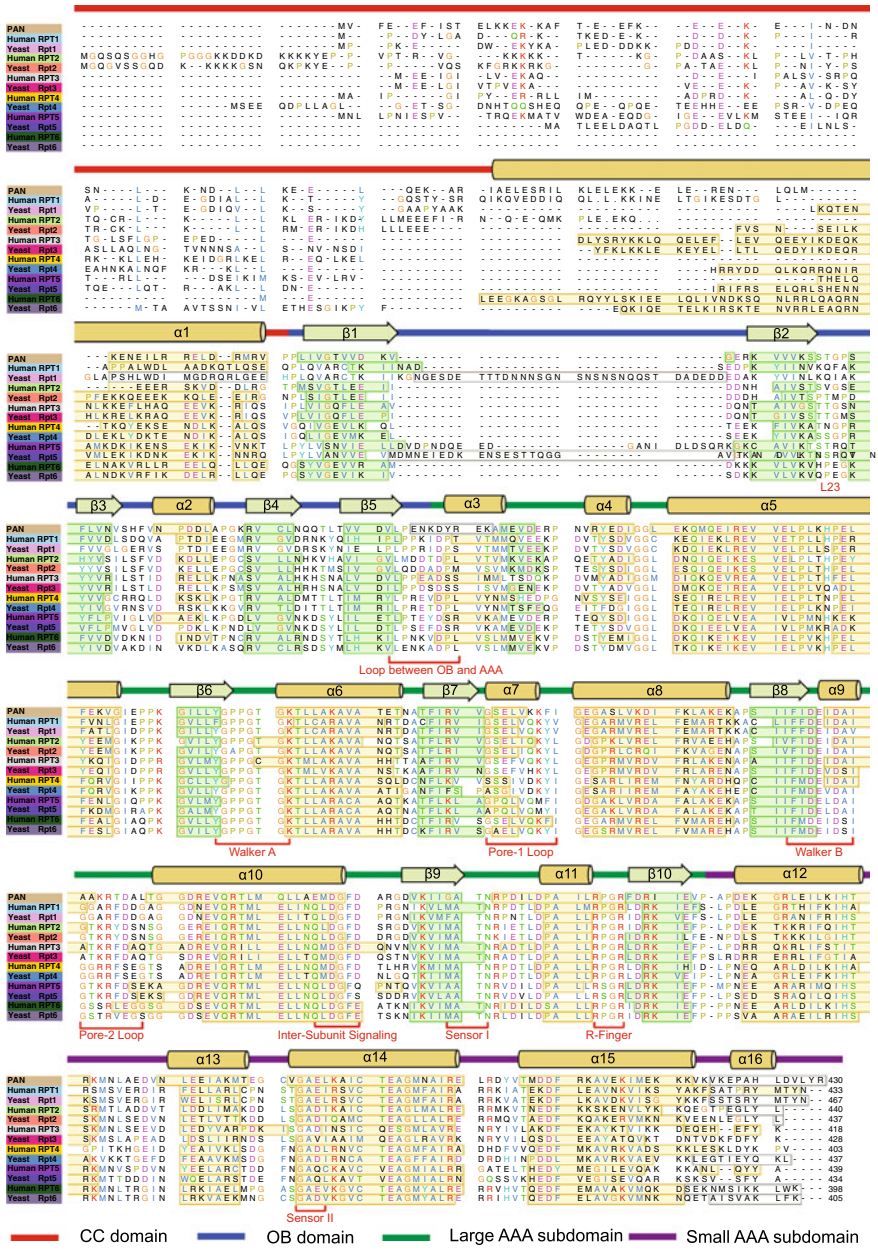


Fig. 1.21 Structure-based sequence alignment. Sequences from the yeast and human proteasomal ATPases are aligned with each other and with that of the archaeal PAN ATPase, based on their structures. The structure of human proteasomal ATPase in state E_A is used as the reference for the sequence alignment

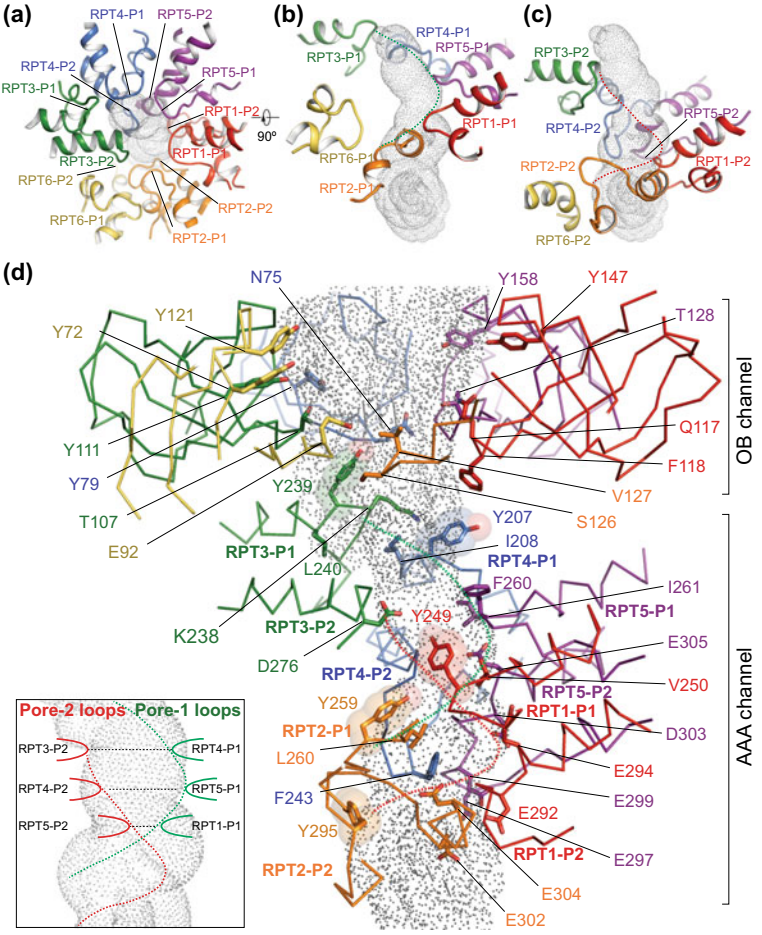


Fig. 1.22 Architecture of the substrate-translocation channel in the human proteasome at a resting state (PDB ID: 6MSB) (Chen et al. 2016a; Dong et al. 2019). **a** Top view of the pore loops aligning along the channel axis from the perspective of the OB domain. The solvent-accessible surface of the channel calculated by the HOLE program is rendered by surface dots. **b** Side views of the pore-1 loops from six RPT subunits decorating the channel, which align along the channel in a spiral staircase formed from RPT1-RPT5, with a backward recession in RPT6 pore-1 loop that is slightly away from the major channel pathway. The pore-1 loops form a helical part of the channel interior as illustrated by the dashed green line. **c** Side view of the pore-2 loops from six RPT subunits decorating the channel, which form another spiral staircase. These pore-2 loops form another helical part of the channel interior, illustrated by the red dashed line. **d** Side view of the complete ATPase channel including components from both the OB and AAA domains, calculated by the HOLE program (Smart et al. 1996). Side-chain patterns observed along the substrate-translocation pathway are highlighted. The five tyrosine residues, highlighted by transparent sphere representation, and a number of hydrophobic and negatively charged residues decorate the AAA channel; color codes for ATPase protomers match those shown in (**a**). Lower left inset, a schematic cartoon showing that the pore-2 loops of RPT3, RPT4, RPT5 pair laterally with the pore-1 loops of RPT4, RPT5 and RPT1, respectively, to form the three narrowest constrictions in the AAA channel in the resting state of the proteasome

channel into the activation of ATP hydrolysis, by modulating the distance between the γ-phosphate of ATP and the Walker B motif (Zhang and Wigley 2008). In contrast to the hydrophobicity of the pore-1 loops, the pore-2 loops are heavily decorated with negatively charged residues, including seven glutamates and two aspartates. The markedly contrasting side-chain patterns in the two complementary parts of the channel surfaces indicate that they play somewhat different roles in substrate translocation.

The AAA channel exhibits three constrictions in the resting state of the proteasome, featuring the positional overlap along the channel axis between the two pore-loop staircases. The aromatic and hydrophobic residues in the pore-1 loops from RPT4, RPT5, and RPT1 are equilaterally paired with adjacent charged residues in the pore-2 loops of RPT3, RPT4 and RPT5, respectively (Fig. 1.22d). Interestingly, mutations in the pore loops resulted in similar phenotypic effects between each pore loop pair. For pore-1 loop, RPT4 has the strongest phenotype, whereas for pore-2 loop the partner RPT3 has the strongest phenotype (Beckwith et al. 2013). The strength of phenotype is also similarly reduced with their distance along the pore axis. In the resting state, the channel radius is constricted to as little as ~2 Å, which must be opened up to allow the insertion of a substrate (Dong et al. 2019). The pore loops of RPT6 is uniquely displaced from the channel axis and does not contribute to these constrictions. Consistently, the pore-2 loop mutations of RPT6 were shown to have little impact on substrate degradation rates (Beckwith et al. 2013).

Dynamic Substrate Interactions with the AAA Ring

The AAA channel formed by the pore loops is structurally plastic (Eisele et al. 2018; Zhu et al. 2018). Even without a substrate, the pore-loop architecture in the states with an open CP are dramatically reorganized between super-helical staircase and saddle-shaped circles deviating from a rigorous staircase arrangement, as typically characterized in state S_{D1}, S_{D2} and S_{D3} (Zhu et al. 2018). While the substrate-free conformation of the ATPase ring in state S_{D2} resembles that of the substrate-bound one in state E_{D2}, the pore-loop conformation is relaxed and less organized in the absence of the substrate (Zhu et al. 2018; Dong et al. 2019). The plasticity of the pore-loop conformation allows the AAA channel to not only accommodate a vast diversity of substrate sequences, but also a hairpin-like polypeptide loop as typically observed in the cryo-EM structure of AAA+ ATPase Vps4 bound with a circular substrate (Han et al. 2019).

In state E_B, the substrate density is approximately centered inside the OB ring (Dong et al. 2019). By contrast, the substrate closely approaches Phe118 of RPT1 inside the OB ring in states E_{C1}, E_{C2}, E_{D1} and E_{D2}. Within the AAA channel, the substrate is gripped by the aromatic residues of pore-1 loops arranging into a right-handed spiral staircase (Fig. 1.23). Either tyrosine or phenylalanine in the pore-1 loops intercalates with the fully stretched main chain of the substrate through hydrophobic interactions. The main chains of the pore-1 loops potentially form hydrogen bonds with the substrate main chain. The adjacent pore-1 loops are spaced

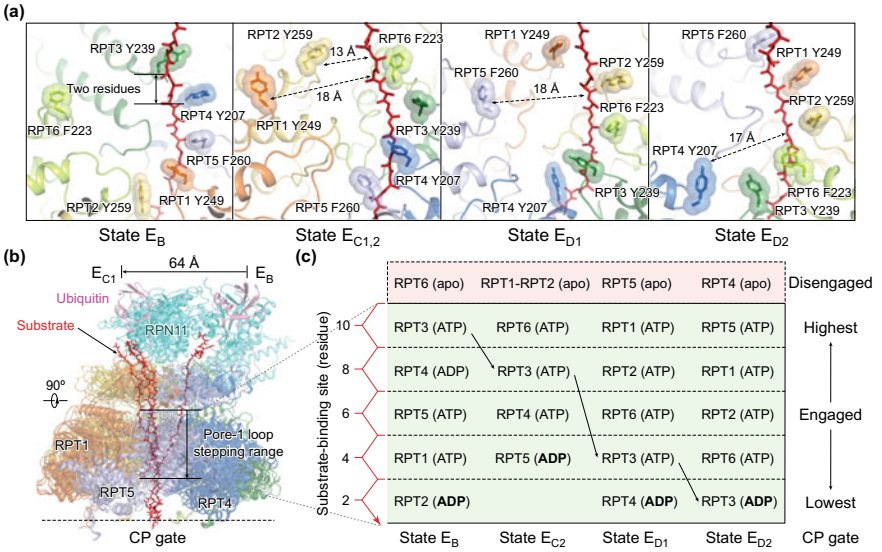


Fig. 1.23 Dynamic substrate-pore loop interactions (PDB IDs: 6MSE, 6MSG, 6MSH, 6MSJ, 6MSK) (Dong et al. 2019). **a** Varying architecture of pore-1 loop staircase interacting with the substrate in five states (E_B, $E_{C1,2}$ and $E_{D1,2}$). Aromatic residues in the pore-1 loops are labeled and shown in stick representation superimposed with transparent sphere representation for highlighting. The distances from disengaged pore-1 loops to the substrate are marked. **b** Side view of all substrate-bound ATPase-RPN11 structures superimposed together based on structural alignment against the CP. **c** A diagram summarizing the axial stepping of the substrate-contacting pore-1 loops and their correlations with nucleotide states. The vertical axis shows the relative location of pore-1 loops interacting the substrate, with the CP positioned at the bottom. The relative distance from the lowest substrate-pore loop contact is labeled using the number of residues as a metric. State E_{C1} is omitted here as its AAA-ATPase structure is identical to that of E_{C2}. The color code of subunits used in all panels is shown in the upper right inset

by two amino acid residues of the substrate, thus evenly distributed along the substrate (de la Pena et al. 2018; Dong et al. 2019). This "two-residue spacing" appears to be a highly conserved structural "rule" for substrate-pore loop interactions among many other AAA+ ATPases, such as p97/Cdc48 (Ripstein et al. 2017; Olszewski et al. 2019), FtsH-like AAA proteases (Puchades et al. 2017, 2019) and Hsp104 disaggregase (Gates et al. 2017). This perhaps implies a common mechanism for the force generation by nonspecific, intercalated stacking interactions between the pore loop's aromatic side chains and the stem of substrate side chains.

The architectural organizations of the substrate-pore loop staircase from states E_B to E_{D2} are all quite comparable, despite different subunit combination. The highest positions in contact with the substrate in states E_B, E_C, E_{D1} and E_{D2} are occupied by the pore-1 loops of RPT3, RPT6, RPT1 and RPT5, respectively. Notably, the pore-1 loop of RPT3 is translated from one end to the other end in the substrate-pore loop staircase during state transitions from E_B to E_{D2}. Meanwhile, the pore-2 loops form another shorter staircase and support the substrate from the opposite of the pore-1

loop staircase (Chen et al. 2016a; Zhu et al. 2018; de la Pena et al. 2018; Dong et al. 2019). In states E_B, E_{D1} and E_{D2}, the pore loops of only one ATPase subunit (RPT6, RPT5 and RPT4, respectively) are dissociated from the substrate, whereas the pore loops of two ATPase subunits (RPT1 and RPT2) are displaced from the substrate in states E_{C1} and E_{C2}. Thus, asymmetric substrate interactions with the pore loops would inevitably exert an allosteric impact on the overall asymmetric ATPase motor conformation. Consistently, functional asymmetry of the six ATPases in substrate degradation has been observed in mutagenesis experiments (Beckwith et al. 2013; Tian et al. 2011; Erales et al. 2012).

Nucleotide-Binding Sites

Three distinct nucleotide states in the proteasomal ATPase subunits have been observed, i.e., ATP-bound, ADP-bound, and apo-like states. The apo-like state is referred to the absence or a very weak density of nucleotide in the nucleotide-binding pocket (Dong et al. 2019). The nucleotide-binding site is located at the Walker A motif located near a short loop between the small and large AAA subdomains (Fig. 1.24). The nucleotide state controls the ATPase conformation by modifying the inter-domain and inter-subunit interactions (Sauer and Baker 2011; Ogura and Wilkinson 2001; Sledz et al. 2013). The proteasome hydrolyzes ATP slowly in the absence of substrate, on average only 27 ATPs per minute per complex (Hoffman and Rechsteiner 1996; Kim et al. 2015; Smith et al. 2011; Benaroudj et al. 2003). ATP hydrolysis in RPT subunits requires magnesium ion bound to the β- and γ-phosphates. The pre-engaged proteasome, despite being saturated with ATP, stays in a basal resting state of minimal ATPase activity (Benaroudj et al. 2003; Peth et al. 2013b; Smith et al. 2011; Kim et al. 2015). Substrate engagement stimulates ATP hydrolysis by tightening the overall ATPase ring structures, by initiating highly concerted inter-subunit interactions and by coordinating conformational changes around the ring (Dong et al. 2019).

The nucleotide states of RPT subunits in the substrate-engaged proteasome exhibit common features reflecting a marked spatiotemporal continuity of the AAA-ATPase motor (Fig. 1.24). Comparison of the pre-engaged structure with the substrate-engaged ones suggests that ATP hydrolysis occurs first in RPT6, followed by hydrolytic events in RPT5 (Dong et al. 2019; Schweitzer et al. 2016; Huang et al. 2016; Chen et al. 2016a). Notably, the ADP-bound states navigate counterclockwise sequentially from RPT6 to RPT3 throughout all six ATPase subunits, indicating a full cycle of coordinated ATP hydrolysis around the AAA-ATPase ring from state E_A to E_{D2}. Magnesium ion density is evident next to ATP and, in fewer cases, also to ADP as well in the cryo-EM density maps of several states (Dong et al. 2019). Whenever the two adjacent ATPase subunits are both in contact with the substrate, their intersubunit interface is stabilized by the insertion of a phenylalanine residue from the ISS motif of the clockwise subunit into a phenylalanine-containing pocket of the counterclockwise subunit near its nucleotide-binding site and pore-1 loop

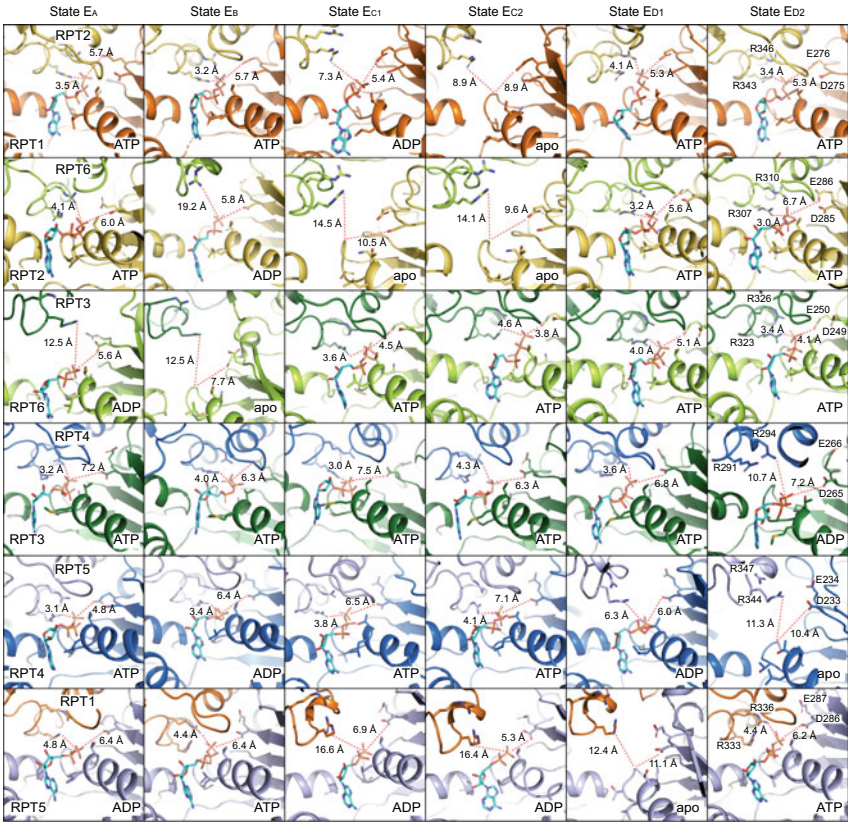


Fig. 1.24 Nucleotide-binding sites in the substrate-bound human proteasome in distinct states (PDB IDs: 6MSB, 6MSD, 6MSE, 6MSG, 6MSH, 6MSJ, 6MSK) (Dong et al. 2019). Comparison of the nucleotide-binding pockets of six ATPases in all states illustrates a common pattern in the geometry of the nucleotide-binding sites. Each row shows the geometry of the nucleotide-binding pocket of one ATPase in all six states. In each panel showing an ATP or ADP-bound state, one red dashed line marks the distance from β/γ-phosphate of nucleotide to the arginine finger of the adjacent ATPase, while the other line marks the distance from the same phosphate to the Walker B motif. In the case of apo-like states, the red lines extend to the proline of the Walker A motif rather than to the phosphate groups. These geometries indicate the potential reactivity of these sites (Zhang and Wigley 2008). When the ATPase is positioned in the middle of pore-loop staircase, but not at the lowest position, the nucleotide-binding pockets are tightly packed whenever ATP or ADP is bound. By contrast, when the ATPase is either in the lowest position of the substrate-pore loop staircase or disengaged from the substrate, the nucleotide-binding pocket is open regardless of whether it is ADP-bound or free of nucleotide. The resolution of 2.8–3.6 Å in the cryo-EM maps of substrate-bound human proteasome allows unambiguous assignment of nucleotide densities to ADP and ATP in the nucleotide-binding pockets of the ATPases. Except for E_A, at least one ATPase subunit in each conformational state exhibits a very weak or partial density for a potential nucleotide in its nucleotide-binding pocket, which precludes de novo atomic modeling of nucleotide coordinates into the density of that subunit. The nucleotide state of these ATPases has been referred to as an apo-like state. It is not possible to differentiate between ATP and ATPγS at the present resolution. Because the mixture of both in the particle population, the 3D classification is not supposed to differentiate ATP against ATPγS at the present resolution. For simplicity, the potential nonhydrolyzed nucleotide has been modelled with ATP

(Chang et al. 2017; Puchades et al. 2017; Dong et al. 2019). This ISS-mediated interaction locks the substrate-bound ATPase subunits together into an approximate rigid body, a mechanical property that is likely required to sustain the translocating force applied by the pore loops on the substrate. It remains to be clarified how intersubunit communications by the ISS motif help to order the timing of ATP hydrolytic events around the ring during substrate translocation.

Substrate Interactions Coupled with ATP Hydrolysis

Key structural motifs of each ATPase subunit couple its interactions with the substrate with the ATP hydrolytic cycle (Ogura and Wilkinson 2001; Lupas and Martin 2002; Iyer et al. 2004; Hanson and Whiteheart 2005; Erzberger and Berger 2006; Sauer and Baker 2011). In other words, the nucleotide states determine the geometry of the nucleotide-binding pockets and are strongly coupled with the substrate-pore loop interactions (Dong et al. 2019). This coupling is reflected in several lines of structural consensus. First, the apo-like state is always observed in the substrate-disengaged ATPases. All apo-like subunits form prominent gaps at their inter-subunit interfaces with their nearest neighbors on both sides. Specifically, the ISS motifs of these apo-like subunits are in a retracted conformation and far away from its adjacent subunit. Likewise, the arginine fingers from the adjacent subunit fall more than 10 Å apart from the Walker A motif of the apo-like subunits, leaving an apparently open nucleotide-binding pocket. Second, the substrate-engaged ATPase that is closest to the CP is always found to be bound with ADP. The clockwise nearest neighbor of the ADP-bound subunit is always found to be in an apo-like state. Third, whenever the substrate-engaged ATPases resides in the middle or top registry in the substrate-pore loop staircase, the nucleotide-binding pocket is always closed by the arginine fingers of the adjacent ATPase. Except states E_A and E_B, these closed nucleotide-binding sites are always occupied by ATP.

The AAA domain remains most stable when bound to ATP. Under this circumstance, the entire AAA domain may still rotate as a rigid body. Systematic structural comparison has revealed that ADP release is associated with the largest hinge-like inter-domain rotation of 15–25° between its small and large AAA subdomains (Fig. 1.25) (Dong et al. 2019). Release of γ-phosphate after ATP hydrolysis does not immediately trigger inter-domain motion of the AAA domain, as the ATPase conformation is also regulated by the inter-subunit interactions. The subsequent disengagement of the ATPase from the substrate facilitates the ADP release, liberates the potential energy that is converted into kinetic energy driving the hinge-like inter-domain rotation. Dissociation from the substrate also allows the ATPase to flip outward by 30–40° relative to the entire ATPase ring. The kinetic energy is liberated and spread out to drive rigid-body rotation of four or five substrate-bound ATPase subunits, propelling the substrate toward the CP (de la Pena et al. 2018; Dong et al. 2019). These structural findings are in line with a sequential model of ATP hydrolysis around the ATPase ring (Smith et al. 2011; Kim et al. 2015; Olivares et al. 2018). Notably, a common feature observed in most other substrate-bound ATPase hexamer

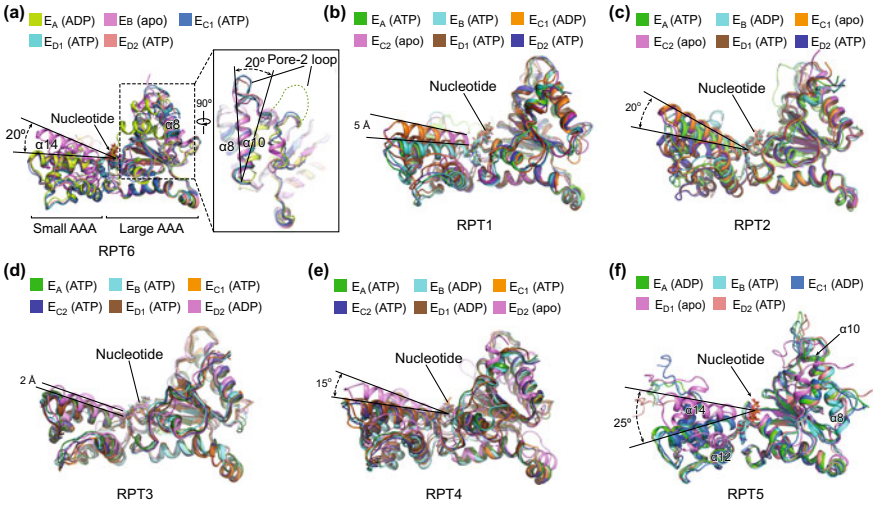


Fig. 1.25 Asymmetric intramolecular dynamics of AAA domains driven by nucleotide exchange (PDB IDs: 6MSB, 6MSD, 6MSE, 6MSG, 6MSH, 6MSJ, 6MSK) (Dong et al. 2019). Superposition of the AAA domain conformations of the same subunit RPT6 (**a**), RPT1 (**b**), RPT2 (**c**), RPT3 (**d**), RPT4 (**e**) or RPT5 (**f**) from distinct states aligned against their large AAA subdomains. Superposition of the RPT6 AAA domain structures from five distinct states aligned against the large AAA subdomain shows that RPT6 assumes three major conformations. As shown in the boxed inset, transition from E_A to E_B involves both refolding of the pore-2 loop, shown in the right insert, and a 20° rigid-body rotation between the large and small AAA subdomains. RPT1, RPT4, and RPT5 assume two major conformations between apo-like and nucleotide-bound states, while RPT2 assumes three major conformations

structures is that at least one subunit is dissociated from the substrate for nucleotide exchange (Gates et al. 2017; Puchades et al. 2017; Monroe et al. 2017; Ripstein et al. 2017; Deville et al. 2017; Alfieri et al. 2018; Thomsen and Berger 2009; Han et al. 2017; Puchades et al. 2019).

Principles of Coordinated ATP Hydrolysis in Functional Regulation

The proteasomal ATPase motor bears a greater degree of inter-subcomplex interactions than most of other AAA+ ATPase motors. On one side, the ATPase motor interfaces with many RPN subunits in the lid and base. On the other side, it forms a highly dynamic, multivalent interface with the α-ring. Such a structural complexity is presumably evolved to achieve the functional complexity of the proteasome. Unexpectedly, three distinct modes of coordinated ATP hydrolysis in the proteasomal ATPase ring appear to regulate the intermediate functional steps of the proteasome (Fig. 1.26) (Dong et al. 2019). The capability of functioning in multiple modes

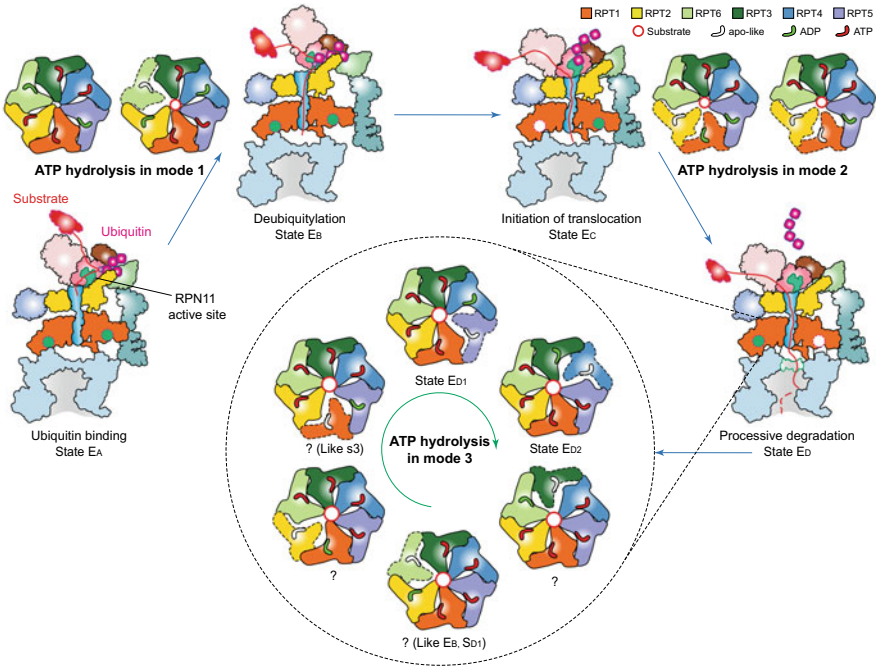


Fig. 1.26 Principal modes of coordinated ATP hydrolysis in the functional regulation of the proteasome (Dong et al. 2019). The cartoon summarizes the concept of three principal modes of coordinated ATP hydrolysis observed in the seven states and the proposed model of how they regulate the complete cycle of substrate processing by the proteasome holoenzyme. Coordinated ATP hydrolysis in *Modes 1*, *2* and *3* features hydrolytic events in two oppositely positioned ATPases, in two consecutive ATPases, and in only one ATPase at a time, respectively. Substrate processing undergoes three major steps before CP gate opening for processive translocation: (1) ubiquitin recognition; (2) simultaneous deubiquitylation and substrate engagement with the AAA-ATPase ring; and (3) translocation initiation, which involves multiple simultaneous events, including ubiquitin release, ATPase repositioning and switching of RPT C-tail insertion pattern. In some cases, the initiation of translocation may precede deubiquitylation. In steps 1 and 2, the ATPases follow the *Mode-1* ATP hydrolysis. In step 3, they follow the *Mode-2* ATP hydrolysis. After the gate is open, the AAA-ATPases hydrolyze ATP in *Mode 3*, in which only one nucleotide is hydrolyzed at a time

implies the existence of multiple feasible pathways of conformational transitions in the same AAA-ATPase hexamer induced by coordinated ATP hydrolysis.

Mode 1 Regulates Ubiquitin Recognition, Initial Substrate Engagement and Deubiquitylation

Mode 1 is observed in states E_{A1}, E_{A2} and E_B and features coordinated ATP hydrolysis in a pair of oppositely positioned ATPases, which is associated with the steps of initial ubiquitin recognition and deubiquitylation (Peth et al. 2010; Worden et al. 2017;

Dong et al. 2019). Before the proteasome gets ready to cleave the ubiquitin chain conjugated to the substrate, the ADP in RPT6 is released and the ATP in both RPT2 and its opposite subunit RPT4 are hydrolyzed. These events trigger refolding of a segment spanning residues 251–266 in RPT6 and drive a marked outward rotation of the entire RPT6 AAA domain (Fig. 1.25). Regulated by this key conformational change, the coordinated ATP hydrolysis in RPT5, and then in RPT4, is expected to relax the conformational rigidity of the AAA ring to enable an iris-like movement opening the AAA channel for substrate insertion (Fig. 1.27). The position of RPT6 at the seam of the AAA ring endows it with certain energetic advantage in allosterically

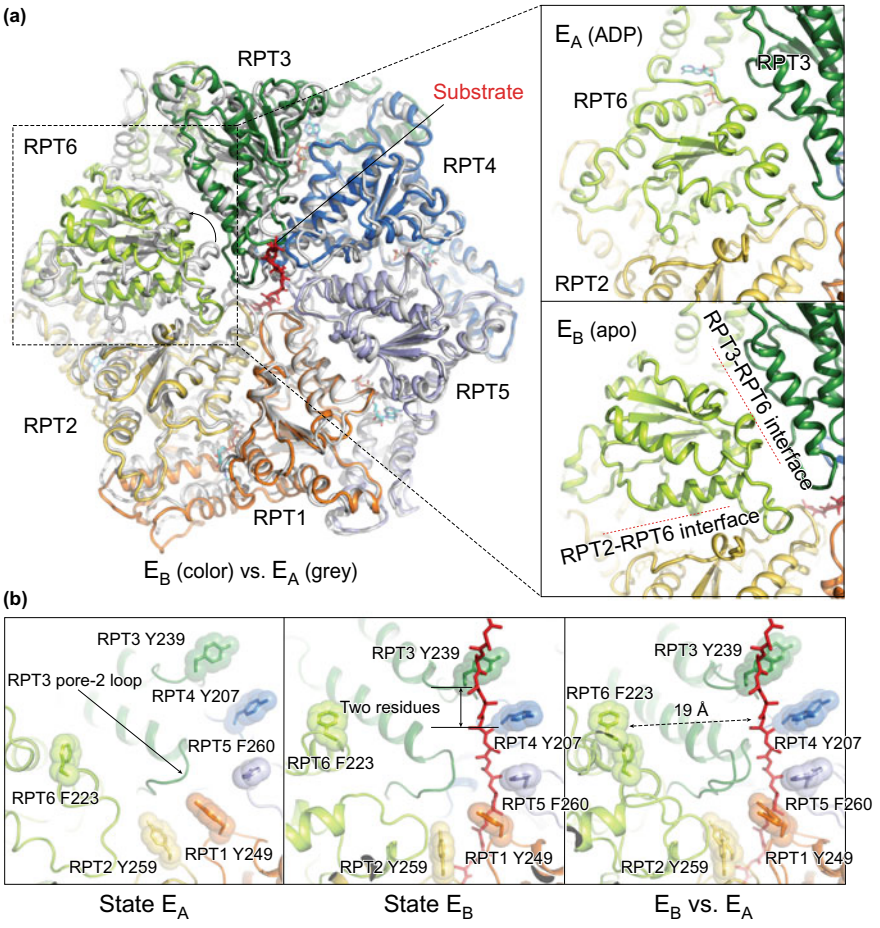


Fig. 1.27 Mechanism of initial substrate engagement with the proteasome (PDB IDs: 6MSB, 6MSD, 6MSE) (Dong et al. 2019). **a** Superposition of the AAA-ring structures of states E_A (grey) and E_B (color). The insets show side-by-side comparison of RPT6 conformations in the four most distant states. Interfacial gaps are marked by red dashed lines. **b** Pore-1 loop staircase before and after engaging with the substrate in state E_B as compared to that in state E_{A1}

triggering the AAA channel opening with little perturbation to the pore loop staircase structure, which is poised to accept an unfolded segment of the substrate without changing the overall ATPase architecture. This mode of coordinated ATP hydrolysis is reminiscent of the nucleotide states in the crystal structure of the ClpX protease, which drive different structural changes in the ATPase ring likely because of the absence of a substrate (Glynn et al. 2009).

Mode 2 Regulates CP Gating, Ubiquitin Release and Initiation of Substrate Translocation

Mode 2 is observed in states E_{C1} and E_{C2} and is characteristic of coordinated ATP hydrolysis in at least two adjacent substrate-disengaged ATPases, which is linked to the intermediate steps of CP gating, ubiquitin release and initiation of substrate unfolding (Dong et al. 2019). Upon deubiquitylation, the proteasome must prepare for substrate translocation and allosteric regulation of the CP gate opening. To this end, ATP hydrolysis in two adjacent subunits, RPT1 and RPT5, is coupled with disengagement of RPT2 and RPT1 from the substrate. As RPT6 acquires ATP again and re-binds the substrate at the top of the substrate-pore loop staircase, forward translocation of the substrate over a two-residue distance is driven by the conformational changes of ATPases from state E_B to E_C. During the following E_{C2}-to-E_{D1} transition, RPT1 and RPT2 both need to re-bind ATP and the substrate and return to the top of the substrate-pore loop staircase, whereas RPT5 is about to release its ADP. As multiple events occur during this process, it is anticipated that several key intermediates were missed in cryo-EM analysis and it remains to be addressed how the E_{C2}-to-E_{D1} transition is accomplished. Recent cryo-EM structures of the yeast Cdc48 and of FtsH-like mitochondrial protease AFG3L2 exhibited a substrate-bound ATPase architecture highly comparable to *Mode 2* of the 26S proteasome (Twomey et al. 2019; Puchades et al. 2019).

Mode 3 Regulates Processive Substrate Unfolding, Translocation and Degradation

Mode 3 is observed in states E_{D1} and E_{D2} and features ATP hydrolysis in only one ATPase at a time during processive substrate unfolding and translocation (Fig. 1.28) (Dong et al. 2019). When the pore-1 loop of the substrate-engaged ATPase reaches the CP-proximal bottom of the substrate-pore loop staircase, the nucleotide state of the ATPase subunit is always ADP-bound. During the process of ADP release, the inter-domain rotation within the AAA domain flips the AAA domain outwards away from the rest of the ATPase ring and detaches its pore loops from the substrate. Meanwhile, its counterclockwise adjacent substrate-engaged ATPase is pushed to the bottom of the ATPase ring, whereas its clockwise neighboring ATPase, which was an apo-like detached seam earlier, now re-binds ATP and the substrate at the top of the ATPase staircase. In concert with these motions, the other three substrate-engaged,

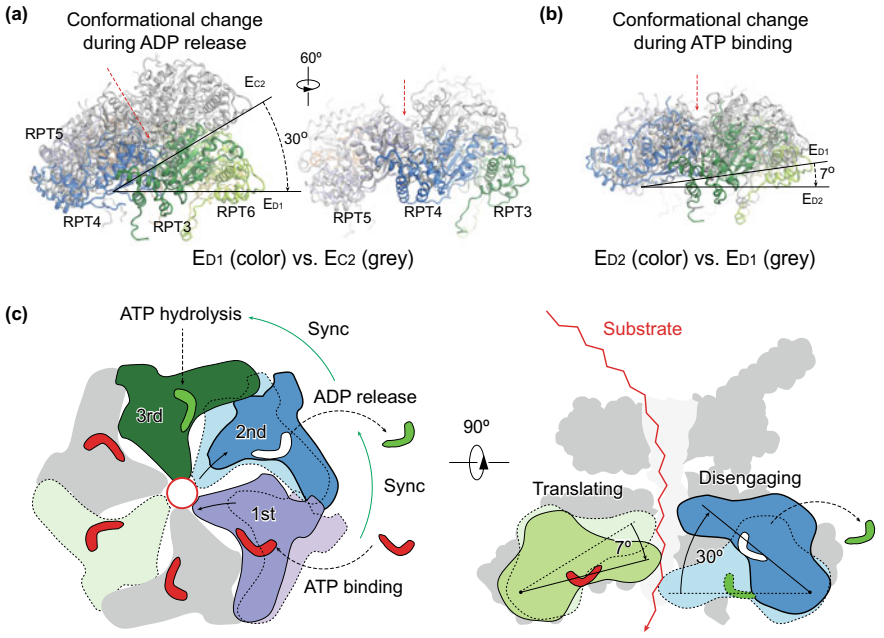


Fig. 1.28 Mechanism of processive substrate translocation in the proteasome (PDB IDs: 6MSH, 6MSJ, 6MSK) (Dong et al. 2019). **a** Structural comparison between E_{D1} (color) and E_{C2} (grey), by using the large AAA subdomain of RPT5 to align the AAA ring structures in distinct states. A 30° out-of-plane rotation is observed during RPT5 disengagement from substrate. The right panel, viewed from a perspective rotated vertically against the left panel, shows that out-of-plane rotation in RPT5 is more prominently amplified in its counterclockwise neighboring ATPases more than its clockwise neighbors. Red arrows mark the center of the AAA ring. **b** Structural comparison between E_{D2} (color) and E_{D1} (grey) in which the large AAA subdomain of RPT5 is used to align the two AAA-ring structures together. A 7° out-of-plane rotation of RPT5 is observed during RPT5 re-association with the substrate. **c** Schematic illustrates the mechanism for processive substrate translocation that involves synchronization of nucleotide processing in three adjacent ATPases, i.e., ATP binding, ADP release and ATP hydrolysis (left), creating vertical rotations in the ATPases that cooperatively propel the substrate (right)

ATP-bound ATPases rotate downwards approximately as a rigid body driven by the intrinsic conformational changes in the substrate-disengaged ATPase undergoing nucleotide exchange. In support of coordinated ATP hydrolysis of *Mode 3* in both human and yeast proteasome (de la Pena et al. 2018; Dong et al. 2019), five distinct conformations of the PAN ATPase ring in the archaeal PAN-CP complex in the absence of a substrate all exhibits only one ATPase disengaged from the rest of the ATPase ring, a key feature consistent with *Mode 3* (Majumder et al. 2019). Both the *Mode-2* and *Mode-3* hydrolysis models confer a unidirectional propagation of conformational changes in the ATPase ring.

The aforementioned analysis postulates on a force generation model integrating various structural observations that seem to suggest a high degree of conservation

in the operating principle of many AAA-ATPase motors, unfoldases, translocases and disaggregates (Fig. 1.28). To generate processive translocation of a substrate, three adjacent ATPases appear to synchronize their nucleotide processing altogether. When the first binds an ATP, the second releases an ADP and the third hydrolyzes an ATP (Fig. 1.28). ATP binding and ADP release trigger a hinge-like rotation of the AAA domain in opposite directions. Substrate disengagement from the second ATPase provides its AAA domain with certain degrees of freedom to amplify the inter-domain hinge-like rotations in both the first and second ATPases into a collective power-stroke-like motion. The largest out-of-plane rotation (~30°) occurs in the substrate-disengaging ATPase. The power-stroke-like conformational change facilitates ATP hydrolysis in the counterclockwise neighboring ATPase, by repositioning the arginine fingers coordinating the neighboring ATP (Zhang and Wigley 2008).

Coordinated ATP Hydrolysis Around the ATPase Ring

Because most substrate-bound AAA+ ATPase complexes were solved in only single conformation per biochemical condition at high resolution, the sequential model of coordinated ATP hydrolysis around the ATPase ring remains largely hypothetical (Cooney et al. 2019; Puchades et al. 2017; Puchades et al. 2019; Monroe et al. 2017; Ripstein et al. 2017). In few studies where several coexisting conformations were obtained, there was often lack of unambiguous intrinsic features labeling the time sequence of those conformations (de la Pena et al. 2018; Majumder et al. 2019; Fei et al. 2020; Ripstein et al. 2020). Interestingly, two hypothetical interpretations of very similar cryo-EM structures for the ClpXP complex, one proposing a strictly sequential model and the other arguing a probabilistic model, illustrates the limit and uncertainty of inferring biochemical mechanism solely from fewer structure snapshots that are lack of time labels or insufficiently represented in intermediate states necessary to substantiate the hypothesized mechanism (Tsai and Hill 2020; Fei et al. 2020; Ripstein et al. 2020). This controversy underscores the importance of obtaining cryo-EM structures that contain features labeling the temporal sequence, as well as key intermediates unambiguously distinguishing between contradicting or alternative mechanistic models. Breaking such interpretive limits or reducing inference ambiguity might require more structural snapshots along the path of conformational changes. For instance, the seven cryo-EM structures of the substrate-bound human 26S proteasome obtained under a common buffer condition contain inherent features of ubiquitin and substrate densities verifying the time sequence of the corresponding states along the path of chemical reactions (Table 1.2) (Dong et al. 2019; Zhang and Mao 2020). These structures establish themselves as a spatiotemporal continuum and provide direct evidence for sequential ATP hydrolysis around the proteasomal ATPase ring with a mixture of *Modes 1*, *2* and *3* (Dong et al. 2019). A rigorously sequential *Mode-3* hydrolysis around the ring still awaits further evidence.

It is anticipated that the proteasome ATPase motor and other similar systems should be versatile enough to allow for the coexistence of multiple pathways of

coordinated ATP hydrolysis. In other words, a rigorously sequential *Mode-3* hydrolysis does not necessarily exclude the possible occurrences of ATP hydrolysis that are less sequential or more stochastic. Indeed, recent cryo-EM structures of the yeast substrate-bound Cdc48-Ufd1-Npl4 complex revealed a nearly planar D1 ATPase ring and a spiral-shaped D2 ATPase ring in *Mode 2* (Twomey et al. 2019). However, the use of ADP/BeF_x resulted in a spiral D1 ring with the D2 ring turning into *Mode 3*. Similarly, cryo-EM structures of two FtsH-like mitochondrial proteases, the engineered soluble Yme1 and AFG3L2 complexes, revealed conformations of substrate-bound ATPase ring highly resembling *Modes 3* and *2* of the 26S proteasome, respectively (Puchades et al. 2017; Puchades et al. 2019). Thus, the fate of the ATP hydrolysis pathway in the hexameric ATPase ring is energetically dependent of its interactions with substrates and regulatory proteins, such as chaperones, shuttle proteins or cofactors.

Summary of Mechanism of the Proteasomal AAA-ATPase Motor

The proteasomal AAA-ATPase motor distinguishes itself from other homohexameric ATPase complexes in several unique ways. First, its pre-engaged conformation is poised to start highly coordinated ATP hydrolysis from RPT6. This role of RPT6 is already encoded in the proteasome structure in the resting state. Second, the mode of coordinated ATP hydrolysis is tightly timed to switch for regulating the necessary intermediate steps of substrate processing. Third, the conformation of the ATPase ring during processive substrate translocation favors fewer major conformational states like E_{D1} and E_{D2} other than six major states (Fig. 1.26) that are expected for a rigorous sequential model, indicating a broken symmetry for the kinetic role of each RPT in substrate translocation (Dong et al. 2019; de la Pena et al. 2018). This also prompts the possibility that the sequential hydrolysis model may neither necessarily nor sufficiently account for all translocation activities. To date, the differential appearances of ubiquitin-substrate densities intrinsic to these structures are the only pieces of evidence for labeling the time sequence of the corresponding conformations along the pathway of chemical reactions (Dong et al. 2019). At least three modes of coordinated ATP hydrolysis have been observed to regulate intermediate steps of substrate processing in the functional proteasome (Dong et al. 2019). Remarkably, each mode of coordinated ATP hydrolysis was comparably observed in separate structural snapshots of various AAA+ ATPases under specific biochemical conditions by independent studies (Glynn et al. 2009; Puchades et al. 2019; Cooney et al. 2019; Twomey et al. 2019; Puchades et al. 2017; de la Pena et al. 2018; Yu et al. 2018; Gao et al. 2019; Ripstein et al. 2020; Deville et al. 2019; Puchades et al. 2020; Zhang and Mao 2020). These studies collectively suggest highly conserved dynamic patterns in the structure-function relationships of AAA+ ATPase hexamers that appear to coexist in the proteasome (Dong et al. 2019; Zhang and Mao 2020).

Mechanism of Substrate Degradation in the CP

Structural Mechanism of Substrate Proteolysis

The α- and β-type subunits of the proteasome are typical members in the superfamily of N-terminal nucleophile (Ntn) hydrolases (Brannigan et al. 1995; Dodson and Wlodawer 1998). Consistent with their sequence similarity, the α- and β-subunits exhibit homologous folds (Lowe et al. 1995; Groll et al. 1997). The N-terminal regions (~ 35 residues) of the α-subunits partly fold into a helix (H0) blocking a groove in the β-stranded sandwich. Due to lack of H0, this cleft is left open in the β-subunits, is referred to as S1 ("specificity") pocket and harbors the active site at Thr1 (Groll et al. 1997). A catalytic triad formed at the N-terminal threonine nucleophile constitutes the active site and is activated upon CP maturation by proteolytic removal of a proximal propeptide occluding the active site (Chen and Hochstrasser 1996; Schmidtke et al. 1996; Seemuller et al. 1996; Huber et al. 2016; Li et al. 2016b). The sidechain Oγ of the N-terminal residue Thr1 of the β-subunits provides the catalytic nucleophile attack of the carbonyl carbon in a peptide bond (Fig. 1.29) (Seemuller et al. 1995a; Lowe et al. 1995). Similar to the proposed hydrolytic mechanism in penicillin acylase (Duggleby et al. 1995), the amino group of Thr1 might be the proton acceptor when Thr1 Oγ adds to an electrophilic entre (Lupas et al. 1995; Seemuller et al. 1995a; Lowe et al. 1995; Groll et al. 1997). Crystal structure of the human CP at 1.8 Å identified a chloride ion, which was previously assigned as catalytic water (NUK), in all active sites, which might aid the nucleophilic attack (Fig. 1.29) (Schrader et al. 2016). The properties of the S1 pocket with respect to substrates are also regulated by other residues of the same subunit and of adjacent subunits β2, β3, and β6, respectively (Groll et al. 1997). Conserved residues essential for the catalytic activity at Thr1 are Glu17 and Lys33 (Seemuller et al. 1995a; Lowe et al. 1995; Arendt and Hochstrasser 1997; Heinemeyer et al. 1997; Seemuller et al. 1996). Other conserved residues, including Ser129, Ser169, and Asp166, seem to be required for structural integrity (Lowe et al. 1995) or β-precursor processing (Seemuller et al. 1995b).

Three β-subunits, β1, β2, and β5, in each β-ring of the eukaryotic CP are proteolytically active. Mutating the Thr1 residues of any of the three β-subunits can disrupt the peptidase activities of the CP (Dick et al. 1998). Similar to classic proteases, the binding pockets are formed at the active sites by specific inter-subunit interactions of the catalytic subunit with its neighboring β-subunit (Borissenko and Groll 2007). The proteasome should not be understood as a simple collection of functionally independent, separated proteases or enzymes but an integral, coherent holoenzyme with multifunctional activities allosterically coupled. Its distributed catalytic activities are spatiotemporally coordinated, and only work properly when the CP is fully assembled and activated. Cleavage of a substrate occurs favorably at the peptide bond located at C-terminal of the P1 residue that is acidic, basic, or hydrophobic (aromatic). Thus, the proteasome displays caspase-like, trypsin-like, and chymotrypsin-like peptide hydrolytic activities (Arendt and Hochstrasser 1997; Voges et al. 1999). These types

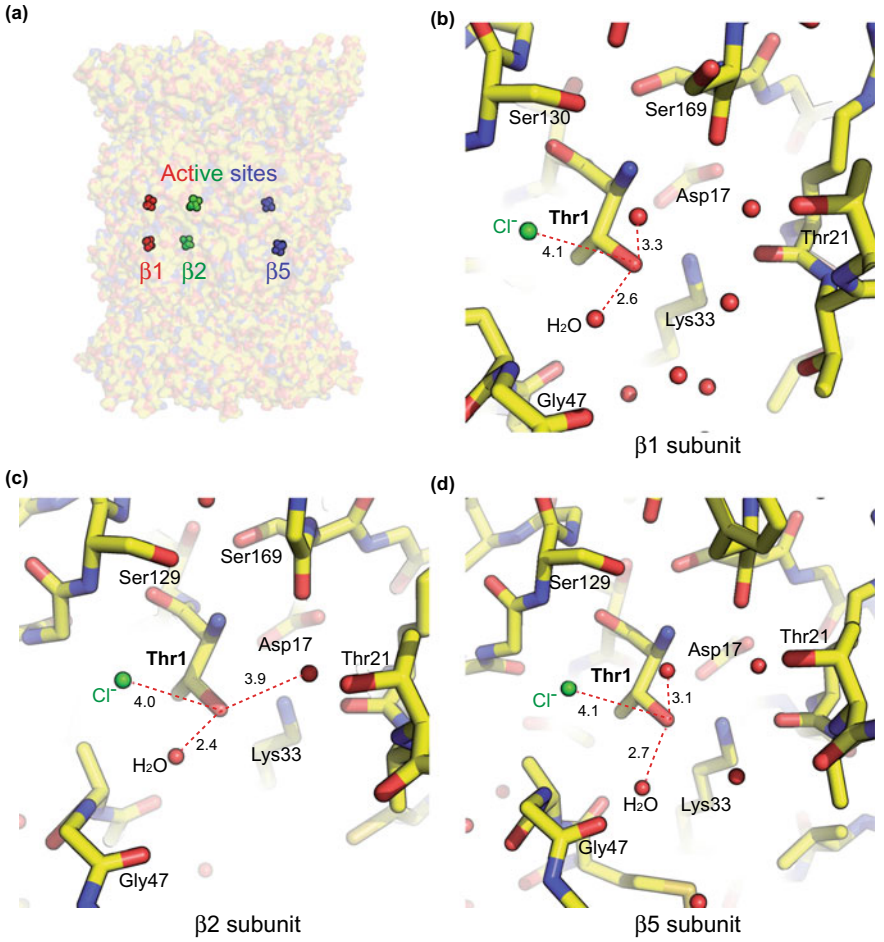


Fig. 1.29 The proteolytically active sites in the 1.8-Å crystal structure of the human CP (PDB ID: 5LE5) (Schrader et al. 2016). **a** Overview of the proteolytically active site of the β-type subunit of the human CP. Thr1 residues in the β1, β2, and β5 subunits are highlighted as sphere representations, whereas the overall CP is rendered as transparent surface. **b**–**d** The close-up view of the residues essential for catalysis around the Thr1 residues in β1 (panel **b**), β2 (panel **c**), and β5 (panel **d**) subunits are shown as stick representations. The chloride ion and water molecules are rendered as green and red spheres, respectively. The distances of the closest water molecules to the side-chain hydroxyl oxygen of Thr1 residue are marked in unit of angstrom

of activity have been thought to associate with specific β1, β2, and β5 subunits, respectively (Wilk and Orlowski 1983; Heinemeyer et al. 1997; Orlowski 1990; Heinemeyer et al. 1991, 1993; Hilt et al. 1993; Enenkel et al. 1994; Dick et al. 1998). Although the three types of active sites exhibit different but weak preference for substrate sequences, they do not have definite sequence specificity and can cleave

peptide bonds at arbitrary positions in an unfolded substrate (Kisselev and Goldberg 2001). Multiple catalytic sites with weak substrate preference offer a unique advantage in concerted processive degradation of diverse cellular proteins by the proteasome (Dick et al. 1991; Bogyo et al. 1998; Wenzel et al. 1994; Ehring et al. 1996; Dolenc et al. 1998; Nussbaum et al. 1998; Arendt and Hochstrasser 1997; Heinemeyer et al. 1997; Groll et al. 1999; Kisselev et al. 1999).

The propeptides of immature β-subunits may play a role in promoting the CP assembly (Chen and Hochstrasser 1996; Mayr et al. 1998; Zuhl et al. 1997). In the newly assembled CP, the propeptides are covalently linked to and bind the Thr1 catalytic sites of the same β-subunits for intramolecular autolysis (Schmidtke et al. 1996; Ditzel et al. 1998). Gly(-1) in the propeptide region is critical for intramolecular processing of the precursor propeptides to activate the β-subunits (Seemuller et al. 1996; Ditzel et al. 1998). Because the catalytic sites also comprise residues of adjacent β-subunits, autolytic activation requires the CP assembly at least in part (Groll et al. 1997). Autolyzed β-subunits (β1, β2, and β5) that are activated may also process the propeptides of other β-subunit precursors (Schmidtke et al. 1996; Ditzel et al. 1998). The propeptides (residues -8 to 1) of the partially processed inactive β-subunits (β3, β6, and β7) are still covalently linked to and shield their Thr1 sites to prevent substrate binding and to inhibit their catalytic activities.

Proteolytic Products by the Proteasome

The CP is thought to degrade substrates in a processive fashion mostly without release of partially degraded intermediates (Dick et al. 1991; Nussbaum et al. 1998; Akopian et al. 1997), although partial degradation by the 26S proteasome has been observed in the activation of many cellular proteins and is likely mediated by the recognition of degradation-stop signal (Saeki et al. 2009a; Zhang et al. 2013; Sandstrom et al. 2019; Chui et al. 2019). The proteolytic chamber efficiently traps unfolded substrates and degrades them into short polypeptides below a certain length limit. The length distribution of released peptide products is broad and ranges from 4 to 25 residues with a length average of 7 to 9 residues (Wenzel et al. 1994; Ehring et al. 1996; Nussbaum et al. 1998; Kisselev et al. 1998). How the CP regulates the product length of degraded substrates has remained enigmatic. One early reasoning attributes to the distance between the active sites that can potentially act in concert (Wenzel et al. 1994). A diffusion-controlled mechanism has been hypothesized (Wenzel et al. 1994; Voges et al. 1999). In the prokaryotic CP, the distance of 2.8 nm between adjacent active sites corresponds to the length of an octa- or nonapeptide in extended conformations (Lowe et al. 1995). However, quantitative analysis of the product length showed a size variation that cannot be reconciled with a solely geometry-based ruler model (Dolenc et al. 1998; Nussbaum et al. 1998; Kisselev et al. 1998). Moreover, these models did not explain why the prokaryotic and eukaryotic proteasomes, which include 14 and 6 active sites, respectively, generate peptides with similar length distributions (Dolenc et al. 1998; Nussbaum et al. 1998; Kisselev et al. 1998).

The possible mechanism of product length regulation by the CP became more comprehensible when more than a dozen non-catalytic substrate-binding sites inside the CP chamber had been observed in the cryo-EM reconstructions of the substrate-engaged human proteasome (Table 1.3) (Dong et al. 2019). Strikingly, all these substrate-binding sites are heavily decorated with aromatic residues, which are also the hallmark of substrate-interacting sites inside the OB and AAA channel. These sites seem to generally stabilize an unfolded peptide segment with a length ranging from 3 to 10 residues and are located in proximity to the three catalytic sites in β1, β2, and β5 (Dong et al. 2019). Their distances to the catalytic sites range from 4 to 25 residues in length, well consistent with the experimental observation of the product length variation. These auxiliary non-catalytic substrate-binding sites may anchor the unstructured substrates during their proteolysis, prohibit refolding of the substrates and prevent the short peptide products from further hydrolytic breakdown. In strong support of this mechanism, long peptides have been observed to degrade faster than short peptides (Dolenc et al. 1998) and NMR studies have shown that the interior of the CP stabilizes unstructured conformations of translocated substrates,

Table 1.3 Substrate interactions with the CP

State	Key contacts at the binding site	Features	Number of substrate residues
E_A, E_B, $E_{C1,2}$, $E_{D1,2}$	Thr1, Cys31 of β2, Cys129 of β3	*C*2 Symmetric	6
E_A	Asn24, Tyr134, Phe137 of β4	At the seam between two β4 subunits	10
E_A	Tyr103 of α1, Tyr61, Tyr90 of β1, and Phe88 of β2	*C*2 Symmetric, at the inter-subunit interface	5
E_A	Asn90 of α5, Tyr90 of β5, Phe101 of β6	*C*2 Symmetric, at the inter-subunit interface	3
E_A, E_B, $E_{C1,2}$, $E_{D1,2}$	Tyr105, Arg117 of α1, His88 of α2	*C*2 Symmetric	3
E_A, E_B, $E_{C1,2}$, $E_{D1,2}$	Tyr59, Cys91, Tyr98 of β4	*C*2 Symmetric	3
E_A	Phe69, Cys91 and Tyr98 of β3	Asymmetric, only present in the chamber with the CP gate open	3
E_A	Ile3, Tyr6, Tyr104 of β3, Tyr120 of β4	Asymmetric, only present in the chamber with the CP gate open	3
E_B, $E_{C1,2}$, $E_{D1,2}$	Tyr30 of β7	*C*2 Symmetric	4

The CP residues that are observed to contact the substrates are recorded and compared, showing a high content of aromatic residues mediating substrate interactions in the interior of the CP. These auxiliary, non-catalytic substrate-binding sites may help regulate the product length distribution of the substrates after proteasomal degradation

which prevents refolding of substrates inside the CP that would sterically hinder their degradation (Ruschak et al. 2010).

Actions of the CP Inhibitors

Structural studies using enzyme inhibitor and site-directed mutagenesis have provided insights into the catalytic mechanism of the CP (Groll and Huber 2004). A crystal structure of the CP inhibited by the peptide aldehyde Ac-Leu-Leu-nLeu-al (ALLN) reveals a hemiacetyl bond with the N-terminal threonine (Thr1) hydroxyl groups of β-subunits (Groll and Huber 2004; Vinitsky et al. 1992; Rock et al. 1994; Figueiredo-Pereira et al. 1994; Vinitsky et al. 1994; Harding et al. 1995). Indeed, mutation of Thr1 to serine retains proteolytic activity, whereas mutation to alanine completely abolishes the proteolytic activity (Groll and Huber 2004). Covalent attachment of a Streptomyces metabolite lactacystin to Thr1 irreversibly inhibits their proteolytic activity (Fenteany et al. 1995). Lactacystin hydrolysis produces clasto-lactacystin beta-lactone that covalently reacts with the active sites (Dick et al. 1996). Other covalent inhibitors have also been investigated, including peptidyl vinyl sulfones (Bogyo et al. 1997b, 1998), dipeptidyl boronic acids (Mc Cormack et al. 1997; Adams et al. 1998) and epoxyketones (Groll and Huber 2004). Most of the covalent inhibitors share a common mechanism of action via reaction with Thr1 Oγ in the β-subunits (Mc Cormack et al. 1997). These CP inhibitors have been extensively used for dissecting the cellular functions of the proteasomes (Bogyo et al. 1997a; Lee and Goldberg 1998; Fenteany and Schreiber 1998).

Since the CP inhibition is an approved approach for cancer therapy, numerous crystal structures of the inhibitor-bound CP complexes have been determined (Huber and Groll 2012; Schrader et al. 2016). The crystal structure of the immunoproteasome CP demonstrates that the active site architecture of the inducible subtype β-subunits (β1i, β2i, and β5i) differs from that of the constitutive CP (Huber et al. 2012). The variation in specificity largely occurs through changes in the S1 pocket and explains why the inhibitor PR-957 specifically binds the β5i subunit. Since the immunoproteasome represents a minor fraction of the proteasome population and is found predominantly in cells involved in the immune response, it is expected that treatment with immunoproteasome-specific inhibitors could result in reduced toxicity (Huber and Groll 2012). Higher resolution (1.9 Å) crystal structure of the human CP in complex with the inhibitor Oprozomib revealed a ring-shaped electron density in the inhibitor-Thr1 conjugate, which was modeled as a heptagonal 1,4-oxazepane ring structure with the C6-methyl group pointed to the inner, solvent-inaccessible side of the ring (Schrader et al. 2016). Thus, the inhibition reaction is mediated via nucleophile attack by the N-terminal amine of the epoxide β carbon (Schrader et al. 2016).

It is also of great clinical merit in developing inhibitors against the proteasomes of infectious pathogens, such as *Mycobacterium tuberculosis* that causes tuberculosis (Lin et al. 2009), the malaria parasite *Plasmodium falciparum* (Li et al.

2016a) and the kinetoplastid parasites *Trypanosoma cruzi*, *Leishmania donovani* and *Trypanosoma brucei* (Khare et al. 2016). Interestingly, binding of oxathiazole-2-one inhibitors induces conformational changes of the *M. tuberculosis* CP, whereas no comparable conformational changes are induced in the human proteasome (Lin et al. 2009). The comparative studies explain why these compounds exhibit specificity for the pathogen proteasome, which provides an important mechanistic guidance for structure-based design of other proteasome inhibitors targeting parasite pathogens (Li et al. 2016a; Khare et al. 2016). Cryo-EM structure of the inhibitor-bound *P. falciparum* proteasome CP provides information of the active-site conformation that can be used for further refinement of inhibitor design (Li et al. 2016a). Furthermore, NMR studies have demonstrated that CP inhibition can also be achieved by binding near the interface between α and β subunits instead of binding directly to the active sites (Sprangers et al. 2008). The novel structural determination tools in cryo-EM and NMR are expected to offer new approaches of developing proteasome inhibitors (Renaud et al. 2018; Scapin et al. 2018).

Proteasome in Health and Disease

Cancer

The proteasome plays a central role in regulating cellular proteostasis, and is recognized as an important regulator of carcinogenesis (Voutsadakis 2017). The proteasome controls the fate of many short-lived proteins regulating cell cycle and tumor suppressors promoting cycle progression, such as cyclin B1, p21, p27 and p53 (Dietrich et al. 1996; Machiels et al. 1997; Adams et al. 1999; Wu et al. 2000; Shah et al. 2001). Dysregulation of these regulatory proteins is associated with malignancies and carcinogenesis. Most cancer cells survive on a considerable level of proteasome activity and are more susceptible to inhibition of proteasome function than are normal cells (Dulic et al. 1994; Pagano et al. 1995; King et al. 1996; Delic et al. 1998; Orlowski et al. 1998). Many cancer types show aberrant functions in their UPS (Hoeller and Dikic 2009; Chen and Madura 2005). A decreased activity of the proteasome has been found in cancer stem cells compared to the rest of cancer cells (Voutsadakis 2017). Proteasome inhibition disrupts cellular protein homeostasis and cell cycle, and potentially activates programmed cell death and apoptosis. The growth of multiple myeloma cell lines is inhibited by several proteasome inhibitors and exhibits anti-inflammatory and immunosuppressive effects (Muchamuel et al. 2009; Basler et al. 2010; Voorhees and Orlowski 2006). Treatment with proteasome inhibitors has been proven beneficiary for patients with multiple myeloma and mantle cell lymphoma in human clinical trials (Orlowski et al. 2002; Jagannath et al. 2004; Roeten et al. 2018). Upstream of the ubiquitin recognition by the proteasome, many E3 ligases and the ATPase p97/VCP that contribute to substrate selectivity for the proteasome are also implicated in cancers (Tang and Xia 2016; Chapman et al. 2011;

Kimonis et al. 2008), and have been considered as potential anti-cancer drug targets due to their crucial roles in proteostasis, protein quality control and cell viability (Anderson et al. 2015; Vekaria et al. 2016; Magnaghi et al. 2013; Skrott et al. 2017).

In clinical trials, the proteasome inhibitor Bortezomib or Velcade (initially called PS-341) exhibited inhibitory efficacy specifically toward hematologic malignancies (Orlowski et al. 2002; Jagannath et al. 2004). Bortezomib is a reversible peptide boronate inhibitor, which preferentially inhibits the chymotrypsin-like activity of the β5 subunit (Huber and Groll 2012). Bortezomib was approved by the US Food and Drug Administration (FDA) for the treatment of refractory multiple myeloma and mantle cell lymphoma (Ruggeri et al. 2009). It also exhibits certain efficacy in treatment of non-small cell lung cancer and pancreatic cancer (Frankland-Searby and Bhaumik 2012). However, Bortezomib treatment gives rise to the development of drug resistance and causes adverse side effects, including asthenia, pain, peripheral neuropathy, disorders, thrombocytopenia, and gastrointestinal, cardiac and pulmonary disorders (Frankland-Searby and Bhaumik 2012).

Several next-generation proteasome inhibitors have been developed for the treatment of hematologic malignancies and solid tumors (Roeten et al. 2018). Some of these inhibitors also target the chymotrypsin-like activity of the β5 subunit, and exhibit increased chemical stability and proteasome-binding affinities, as well as altered toxicities in clinical trials (Huber and Groll 2012). The US FDA approved the proteasome inhibitor Carfilzomib in 2012 for the treatment of patients with multiple myeloma. In contrast to Bortezomib, Carfilzomib is an irreversible peptide epoxyketone inhibitor of the CP (Wang et al. 2013). However, Carfilzomib has shown efficacy in multiple myeloma patients who developed drug resistance against Bortezomib (Wang et al. 2013).

In addition to the inhibitors targeting the CP, a bis-benzylidine piperidone RA190 and its derivatives like RA183 were developed as covalent inhibitors of RPN13 for the treatment of ovarian cancer (Anchoori et al. 2013, 2018; Kisselev 2013; Lu et al. 2017a) and multiple myeloma (Song et al. 2016, 2019; Trader et al. 2015). NMR structural studies found that RA190-binding site on RPN13 PRU domain coincides with the interaction site of RPN13 with a proline-rich C-terminal extension of RPN2 (Lu et al. 2017a). PA190 is found to also directly bind and inactivate UCH37, causing accumulation of substrates at the proteasome. Thus, it has been proposed that RA190 targets both RPN13 and UCH37 in a parallel fashion to effectively block substrate degradation (Lu et al. 2017a).

Aging

Progressive aggregation of damaged or misfolded proteins is a hallmark of aging cells, and is associated with decline in cellular proteostasis function and UPS impairment, which is also implicated in aging-related disease like Alzheimer's disease (AD) and dementia (Vilchez et al. 2014; Hegde et al. 2019; Fischer et al. 2009; Saez and Vilchez 2014). The ability of the proteasome in maintaining proteostatsis is generally

challenged and prone to decline in aging cells, tissues and organisms, such as human epidermal cells and lymphocytes, bovine eye lens and rat muscle, lung, liver, kidney, hippocampus, spinal cord, cerebral cortex and heart (Baraibar et al. 2012; Saez and Vilchez 2014). Age-related proteasome dysregulation may occur at multiple levels, including suppressed expression of proteasome subunits (Lee et al. 1999), reduced proteolytic activity caused by oxidative damage (Bulteau et al. 2001; Wang et al. 2010), and proteasome disassembly (Wang et al. 2010; Vernace et al. 2007; Bajorek et al. 2003). Part of these is also reflected in altered function in ubiquitin ligases, such as cytoplasmic E3 ligase mahogunin (MGRN1) that is found instead in nucleus in aging hippocampal neurons (Benvegnu et al. 2017), and the reduced level of UBE3A in the hippocampus of elderly rats (Fletcher et al. 2014).

Aggregated proteins might form non-productive, stalled complex with the proteasome, deplete the intracellular proteasome reservoir and thus reduce their overall cellular activity (Grune et al. 2004). When the constitutive β5 subunit was substituted with the thymus-specific subtype β5t subunit of reduced chymotrypsin-like activity, the transgenic mice developed signs of early aging and suffered a much shorter lifespan (Tomaru et al. 2012). Consistently, several studies on long-lived humans (centenarians) and animals, such as the exceptionally long-lived naked mole rat and a long-lived bat species and an exceptionally long-lived invertebrate called giant clam, have found correlates with increased proteasome activity and capacity (Chondrogianni et al. 2000; Perez et al. 2009; Salmon et al. 2009; Ungvari et al. 2013). These data suggest that the overall level of proteasome activity may be correlated with the age of an organism. This hypothesis has been further supported by genetic studies in several model organisms (Chen et al. 2006; Tonoki et al. 2009; Kruegel et al. 2011).

Neurodegenerative Diseases

Aggregation of misfolded proteins compromises cellular proteostasis and is characteristic of neurodegenerative diseases, such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), Amyotrophic lateral sclerosis (ALS) and spongiform encephalopathies (Selkoe 2003; Rubinsztein 2006; Brettschneider et al. 2015). Maintaining appropriate proteostasis in neurons is crucial due to their complex organization, function and prolonged lifespan. Unchecked aggregation of misfolded proteins can impair normal cellular function in the cytoplasm, nucleus or extracellular space. More importantly, neurons are unable to dilute aggregate loads through cell division (Tai and Schuman 2008). These underscore the critical function of the UPS in neuronal synapses. For example, synaptic protein homeostasis, plasticity and long-term memory formation rely on the proteome remodeling that is tightly regulated by the UPS (Fonseca et al. 2006; Tai and Schuman 2008; Aso et al. 2012; Djakovic et al. 2012). The membrane-anchored proteasome modulates neuronal function in the

mammalian nervous system, degrades intracellular proteins and releases the polypeptide products into the synaptic cleft, where they trigger neuronal signaling via postsynaptic N-methyl-D-aspartate receptor for synaptic regulation (Ramachandran and Margolis 2017).

Neurodegenerative diseases are mostly associated with aging, and share many traits in proteostasis, such as impaired UPS function and reduced proteasome activity (Vilchez et al. 2014; Hegde et al. 2019; Fischer et al. 2009; Saez and Vilchez 2014; Keller et al. 2000; McNaught et al. 2002; Ciechanover and Brundin 2003; Rubinsztein 2006; Ortega et al. 2007; Ortega and Lucas 2014). Aggregation of misfolded proteins causes loss of proteostasis and is commonly accompanied with progressive death of neurons (Selkoe 2003; Rubinsztein 2006; Brettschneider et al. 2015). Brain region-specific proteasome inhibition is implicated in the neuropathology and is a clinical hallmark of neurodegenerative diseases (McNaught et al. 2002, 2004; Bedford et al. 2008; Li et al. 2010). In the mice model of AD with mutant amyloid precursor protein (APP) transgene (Tg2576), $A\beta_{1\text{-}42}$ seems to inhibit proteasome activity and DUBs and enhances amyloid and tau accumulation (Almeida et al. 2006; Oh et al. 2005; Tseng et al. 2008), likely by stalling the proteasome in the Aβ aggregation body (Guo et al. 2018). Interestingly, a higher level of immunoproteasome expression was observed in the brain of AD patients compared to the brains of the non-demented elderly (Mishto et al. 2006). Going forward beyond the "Aβ hypothesis" of AD (Morris et al. 2018; Mullane and Williams 2019), it remains to be understood how the proteasome and UPS are related to other factors contributing to neurodegenerative diseases, such as inflammation and insulin resistance in the brain.

Secretory proteins and integral membrane proteins are synthesized and enter the ER lumen for proper protein folding and post-translational modifications. Misfolded proteins in the ER are removed by the pathway of endoplasmic reticulum-associated protein degradation (ERAD), which prevents neurodegenerative diseases (Wu and Rapoport 2018). Impaired ERAD and ER stress are associated with polyglutamine (polyQ) toxicity commonly found in several neurodegenerative disorders (Duennwald and Lindquist 2008; Remondelli and Renna 2017; Smith and Mallucci 2016). There are three distinct ERAD pathways, ERAD-L, ERAD-M, and ERAD-C, in which the misfolded protein domain is localized in the ER lumen, within the membrane, or on the cytosolic side of the ER membrane, respectively (Huyer et al. 2004; Vashist and Ng 2004; Carvalho et al. 2006). Another pathway is used to remove misfolded protein from the inner nuclear membrane (Foresti et al. 2014; Khmelinskii et al. 2014). Through ubiquitylation, the p97/Cdc48 ATPase extracts the substrates from the membrane and delivers them to the proteasome for degradation (Bays et al. 2001; Braun et al. 2002; Jarosch et al. 2002; Rabinovich et al. 2002). Proteasome inhibition impairs ERAD and allows misfolded proteins to aggregate in the ER, which activates the unfolded protein response (UPR) (Tsai and Weissman 2010). UPR regulates the gene expression controlling protein folding and ERAD to restore ER homeostasis (Travers et al. 2000). However, prolonged ER stress that is out of control by UPR activation eventually leads to cell death (Travers et al. 2000). Besides neurodegenerative diseases, many other physiological conditions, such as hypoxia, glucose deprivation, oxidative stress and certain mutations in proteins, can also cause

aggregation of misfolded proteins in the ER and trigger UPR activation (Tsai and Weissman 2010).

Immune Diseases

The proteasome is extensively involved in the regulation of the immune system, including antigen presentation, NF-κB pathway, and NLRP1 inflammasome activation. The proteasome mediates antigenic protein degradation into short peptides that are presented on the MHC-I complex (Ferrington and Gregerson 2012). The NF-κB transcription factors (NF-κB and Rel proteins) regulate gene expression in innate and adaptive immunity, inflammation, stress responses, B-cell development and lymphoid organogenesis. NF-κB is a prosurvival pathway and is upregulated in many inflammatory diseases and cancers (Wang et al. 1996). In cancer cells, NF-κB is involved in the expression of the antiapoptotic IAP family of genes as well as BCL-2 prosurvival genes (Wang et al. 1998; Zong et al. 1999; Chen et al. 2000). The proteasome mediates degradation of regulatory elements for transcriptional activation in both canonical and noncanonical pathways of NF-κB activation. In the inactive state, NF-κB/Rel is bound to and sequestered by the inhibitory IκB subunits in the cytoplasm (Baldwin 2001). In the canonical pathway, proinflammmatory cytokines activate the IκB kinase (IKK) complex that phosphorylates IκB and causes IκB ubiquitiylation for proteasome-mediated degradation that activates the NF-κB/RelA complex (Chen et al. 1995; Scherer et al. 1995; Spencer et al. 1999; Winston et al. 1999). In the noncanonical pathway, the NF-κB-inducing kinase activates IKKα that phosphorylates the C-terminal residues of NF-κB2/p100. Phosphorylated NF-κB is ubiquitylated and partially degraded by the proteasome into transcriptionally active NF-κB2/p52.

Mammalian cells use a diverse spectrum of pattern-recognition receptors (PRRs) to monitor cytoplasmic microbial activities and pathogen invasion for signaling innate immune responses. Several PRRs form megadalton multiprotein complexes named the inflammasome upon recognition of their cognate danger signals. The inflammasome recruits and activates caspase-1, which then cleaves and activates inflammatory cytokines and gasdermin D (GSDMD), and triggers inflammatory cell death called pyroptosis (Zhang et al. 2015; Broz and Dixit 2016; Sharif et al. 2019). NLRP1B is one of the inflammasome-forming PRRs and a crucial member in the nucleotide-binding domain leucine-rich repeat (NLR) protein superfamily. Several studies have found that proteasomal degradation of N-terminal domains of the NLRP1B protein releases the C-terminal fragment of the NLRP1B containing a caspase activation and recruitment domain (CARD) domain, which is sufficient to self-assemble, recruit caspase-1 and active the inflammasome (Sandstrom et al. 2019; Chui et al. 2019). The degradation-dependent activation of inflammasome likely underlies its ability to sense any pathogen effectors that can induce proteasomal degradation of NLRP1B, such as IpaH7.8, an E3 ubiquitin ligase secreted by

the pathogen *Shigella flexneri* that ubiquitylates NLRP1B and lead to inflammasome activation in *Shigella*-infected macrophages (Sandstrom et al. 2019).

Proteasome inhibition can give rise to decreased inflammatory and immune responses, as well as compromised cell migration and adhesion (Frankland-Searby and Bhaumik 2012). These effects are also reflected in the proteasome function in antigen presentation. Lymphopenia is a toxicity effect commonly observed in clinical administration of Bortezomib (Jagannath et al. 2004). Proteasome inhibition by Bortezomib induces suppressive effects on the T cell compartment, potentially leading to T cell apoptosis in monocyte-derived dendritic cells (Nencioni et al. 2006). Activated T cells are more susceptible to proteasome inhibition than resting T cells (Blanco et al. 2011). Proteasome inhibition by Bortezomib can deplete alloreactive T cells in vitro, while preserving immune response against pathogens (Blanco et al. 2011). These observations suggest that proteasome inhibition represents a promising option for the treatment of diseases involving activated T cells. This hypothesis is supported by several preclinical studies in mouse models for autoimmune encepahlomyelitis, arthritis, colitis, graft-versus-host disease, systemic lupus erythematosus, myastenia gravis and allograft rejection (Moran et al. 2012).

Cardiovascular Diseases

The UPS plays an important role in cardiac physiology and disease (Drews and Taegtmeyer 2014; Li and Wang 2011; Pagan et al. 2013; Shukla and Rafiq 2019). Impairment of UPS function may cause a number of cardiac diseases, such as heart failure, cardiomyopathies, hypertrophy, atrophy, ischemia-reperfusion, and atherosclerosis. Increased oxidative damage to proteins, elevated levels of ubiquitylated proteins and proteasome dysfunction have been found in several cardiac diseases, such as hypertrophic, diabetic cardiomyopathies and ischemic heart diseases (Powell et al. 2012; Shukla and Rafiq 2019). Reduced proteasome activity has been linked to myocardial ischemia/reperfusion (I/R) injury (Bulteau et al. 2001; Tian et al. 2012; Li et al. 2011). A transgenic mouse line that overexpresses PA28α in the heart exhibited increased proteasome activity, which protected the animals against I/R injury (Li et al. 2011). In a mouse model of desmin-related cardiomyopathy (DRC), cardiac PA28α overexpression suppressed cardiac hypertrophy and prolonged the animal lifespan (Li et al. 2011). In many forms of cardiomyopathies, such as I/R injury and diabetic cardiomyopathy, inducible immunoproteasome expression is significantly elevated (Cai et al. 2008; Gomes et al. 2006; Powell et al. 2008). Similar to neurological toxicity, proteasome inhibition with Bortezomib seems to be associated with an increase in cardiac dysfunction, such as congestive heart failure or arrythmia (Schlossarek and Carrier 2011). Short-term and low-dose local treatments with proteasome inhibitors have produced contradictive results that await further clarification (Powell et al. 2012). Although both the inhibition and enhancement of proteasome activity can potentially confer cardioprotection under distinct conditions, concerns and controversies

remain for the toxicity of impairing the proteasome in long-term therapies (Pagan et al. 2013; Li and Wang 2011; Shukla and Rafiq 2019).

Perspective

In this chapter, I reviewed nearly five decades of studies on the structures and functions of the proteasome. The past five years have seen a major culmination of breakthrough studies in elucidation of the inner working of the proteasome in both human and yeast. Progressive improvements on the biochemical and cryo-EM methods have yielded atomic-level information critical for understanding the functional dynamics of the proteasome. Not only have the conformational states of the proteasome been extensively mapped, but also their correspondence to the intermediate steps of substrate processing and a nearly complete degradation pathway within the proteasome have been reconstructed at the atomic level. A breathtaking picture of the "central dogma" of the proteasome is now emerging and lays an essential foundation for further understanding how the UPS decodes the ubiquitin signals as well as for the development of proteasome-related therapeutics. Looking forward, a major knowledge gap exists between numerous ubiquitylation pathways and the intrinsic working machineries of the proteasome, the understanding of which poses much greater technical challenges due to the transient nature of the interactions of the proteasome with the extrinsic receptors, DUBs, the ubiquitin chain-elongating E4s, and an expanding network of proteasome-related proteins. Another great challenge is how we can best convert what we know about the proteasome into clinical tools and therapeutics for combating human diseases and improving human wellbeing. With further developments in chemical and biological methods, cryo-EM imaging approaches and data science tools, addressing these challenges will soon become possible in years to come.

Future Problems

- What are the missing key intermediate states necessary for unambiguously defining the detailed mechanisms for the complete cycle of substrate processing in the proteasome? And, how are these intermediate states associated with the proteasome function in vivo?
- Are there alternative pathways of coordinated ATP hydrolysis in the proteasomal AAA-ATPase motor? How are these hydrolytic pathways regulated at the proteasome level?
- How do the intrinsic ubiquitin receptors recognize the polyubiquitin chains in the proteasome? What are the conformational dynamics of RPN13 in the proteasome? What are the unidentified intrinsic ubiquitin or UBL receptors in the proteasome? How do they coordinate with one another? What are their functional differences and specificities in recognizing ubiquitin signals and UBL proteins?
- How do the extrinsic ubiquitin receptors deliver substrates to the proteasome? How do the extrinsic receptors coordinate their ubiquitin recognitions with

the intrinsic receptors? What are the other cellular proteins involved in the delivery of substrates by the extrinsic receptors? Are there any other unidentified extrinsic ubiquitin receptors associated the proteasome? What are the differences in substrate selectivity between the intrinsic and extrinsic ubiquitin receptors?

- How do those non-essential DUBs such as USP14 and UCH37 work with the proteasome? How do their interactions regulate the proteasome structure and function during substrate processing? Are they involved in the process of substrate delivery by the shuttle receptors?
- How do the proteasome-associated E3 and E4 enzymes process substrates in the proteasome? How do their interactions regulate the proteasome structure and function? Are there also other E2s and E3s associated with the proteasome for substrate processing? How do they work together?
- How do the proteasome-associated proteins function in the proteasome in the presence of substrates? How do their interactions with the proteasome regulate the proteasome structure and function?
- How do those post-translational modifications of the proteasome regulate the ways the proteasome processes substrates? What are their cellular functions?
- What are the assembly pathways of the RP in mammalian cells? How does the dynamic assembly of the RP affect or regulate the proteasome function?
- Are there inhibitors of the proteasome specifically interacting with the RP instead of the CP? If any, how do these inhibitors contribute to the improvement of human health?
- What are the exact underlying mechanisms of proteasome inhibitors in vivo? Why is the same proteasome inhibitor beneficiary for certain diseases but detrimental for others?
- How can we best harvest the proteasomes or UPS for combating neurodegenerative diseases, and of course, other closely related human diseases in general, including cancer and immune diseases?

Acknowledgements The author thanks all members of the Mao Laboratory for helpful discussions and particularly Shitao Zou and Shuwen Zhang for a proofreading of the manuscript, Drs. Alfred Goldberg, Daniel Finley, Marc Kirschner, and Ying Lu for constructive discussions relevant to this work. I apologize for not including many relevant excellent literatures due to our current focus and scope of this review and my limitation in referencing efforts. This work was partly funded by National Natural Science Foundation of China, grant number 1177402 and Natural Science Foundation of Beijing Municipality, grant number Z180016.

References

Adams J, Behnke M, Chen S, Cruickshank AA, Dick LR, Grenier L, Klunder JM, Ma YT, Plamondon L, Stein RL (1998) Potent and selective inhibitors of the proteasome: dipeptidyl boronic acids. Bioorg Med Chem Lett 8(4):333–338. https://doi.org/10.1016/s0960-894x(98)00029-8

Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, Lazarus DD, Maas J, Pien CS, Prakash S, Elliott PJ (1999) Proteasome inhibitors: a novel class of potent and effective antitumor agents. Cancer Res 59(11):2615–2622

Adrain C, Creagh EM, Cullen SP, Martin SJ (2004) Caspase-dependent inactivation of proteasome function during programmed cell death in Drosophila and man. J Biol Chem 279(35):36923–36930. https://doi.org/10.1074/jbc.M402638200

Ahn JY, Tanahashi N, Akiyama K, Hisamatsu H, Noda C, Tanaka K, Chung CH, Shibmara N, Willy PJ, Mott JD et al (1995) Primary structures of two homologous subunits of PA28, a gamma-interferon-inducible protein activator of the 20S proteasome. FEBS Lett 366(1):37–42. https://doi.org/10.1016/0014-5793(95)00492-r

Aki M, Shimbara N, Takashina M, Akiyama K, Kagawa S, Tamura T, Tanahashi N, Yoshimura T, Tanaka K, Ichihara A (1994) Interferon-gamma induces different subunit organizations and functional diversity of proteasomes. J Biochem 115(2):257–269. https://doi.org/10.1093/oxfordjournals.jbchem.a124327

Akopian TN, Kisselev AF, Goldberg AL (1997) Processive degradation of proteins and other catalytic properties of the proteasome from Thermoplasma acidophilum. J Biol Chem 272(3):1791–1798. https://doi.org/10.1074/jbc.272.3.1791

Albert S, Schaffer M, Beck F, Mosalaganti S, Asano S, Thomas HF, Plitzko JM, Beck M, Baumeister W, Engel BD (2017) Proteasomes tether to two distinct sites at the nuclear pore complex. Proc Natl Acad Sci U S A 114(52):13726–13731. https://doi.org/10.1073/pnas.1716305114

Alfano C, Faggiano S, Pastore A (2016) The ball and chain of polyubiquitin structures. Trends Biochem Sci 41(4):371–385. https://doi.org/10.1016/j.tibs.2016.01.006

Alfieri C, Chang L, Barford D (2018) Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13. Nature 559(7713):274–278. https://doi.org/10.1038/s41586-018-0281-1

Almeida CG, Takahashi RH, Gouras GK (2006) Beta-amyloid accumulation impairs multivesicular body sorting by inhibiting the ubiquitin-proteasome system. J Neurosci 26(16):4277–4288. https://doi.org/10.1523/JNEUROSCI.5078-05.2006

Ambroggio XI, Rees DC, Deshaies RJ (2004) JAMM: a metalloprotease-like zinc site in the proteasome and signalosome. PLoS Biol 2(1):E2. https://doi.org/10.1371/journal.pbio.0020002

Anchoori RK, Jiang R, Peng S, Soong RS, Algethami A, Rudek MA, Anders N, Hung CF, Chen X, Lu X, Kayode O, Dyba M, Walters KJ, Roden RBS (2018) Covalent Rpn13-Binding inhibitors for the treatment of ovarian cancer. ACS Omega 3(9):11917–11929. https://doi.org/10.1021/acsomega.8b01479

Anchoori RK, Karanam B, Peng S, Wang JW, Jiang R, Tanno T, Orlowski RZ, Matsui W, Zhao M, Rudek MA, Hung CF, Chen X, Walters KJ, Roden RB (2013) A bis-benzylidine piperidone targeting proteasome ubiquitin receptor RPN13/ADRM1 as a therapy for cancer. Cancer Cell 24(6):791–805. https://doi.org/10.1016/j.ccr.2013.11.001

Anderson DJ, Le Moigne R, Djakovic S, Kumar B, Rice J, Wong S, Wang J, Yao B, Valle E, Kiss von Soly S, Madriaga A, Soriano F, Menon MK, Wu ZY, Kampmann M, Chen Y, Weissman JS, Aftab BT, Yakes FM, Shawver L, Zhou HJ, Wustrow D, Rolfe M (2015) Targeting the AAA ATPase p97 as an approach to treat cancer through disruption of protein homeostasis. Cancer Cell 28(5):653–665. https://doi.org/10.1016/j.ccell.2015.10.002

Arcus V (2002) OB-fold domains: a snapshot of the evolution of sequence, structure and function. Curr Opin Struct Biol 12(6):794–801

Arendt CS, Hochstrasser M (1997) Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. Proc Natl Acad Sci U S A 94(14):7156–7161. https://doi.org/10.1073/pnas.94.14.7156

Arrigo AP, Tanaka K, Goldberg AL, Welch WJ (1988) Identity of the 19S 'prosome' particle with the large multifunctional protease complex of mammalian cells (the proteasome). Nature 331(6152):192–194. https://doi.org/10.1038/331192a0

Asai M, Tsukamoto O, Minamino T, Asanuma H, Fujita M, Asano Y, Takahama H, Sasaki H, Higo S, Asakura M, Takashima S, Hori M, Kitakaze M (2009) PKA rapidly enhances proteasome

assembly and activity in in vivo canine hearts. J Mol Cell Cardiol 46(4):452–462. https://doi.org/10.1016/j.yjmcc.2008.11.001

Asano S, Fukuda Y, Beck F, Aufderheide A, Forster F, Danev R, Baumeister W (2015) Proteasomes. A molecular census of 26S proteasomes in intact neurons. Science 347(6220):439–442. https://doi.org/10.1126/science.1261197

Aso E, Lomoio S, Lopez-Gonzalez I, Joda L, Carmona M, Fernandez-Yague N, Moreno J, Juves S, Pujol A, Pamplona R, Portero-Otin M, Martin V, Diaz M, Ferrer I (2012) Amyloid generation and dysfunctional immunoproteasome activation with disease progression in animal model of familial Alzheimer's disease. Brain Pathol 22(5):636–653. https://doi.org/10.1111/j.1750-3639.2011.00560.x

Aufderheide A, Beck F, Stengel F, Hartwig M, Schweitzer A, Pfeifer G, Goldberg AL, Sakata E, Baumeister W, Forster F (2015) Structural characterization of the interaction of Ubp6 with the 26S proteasome. Proc Natl Acad Sci U S A 112(28):8626–8631. https://doi.org/10.1073/pnas.1510449112

Bader M, Benjamin S, Wapinski OL, Smith DM, Goldberg AL, Steller H (2011) A conserved F box regulatory complex controls proteasome activity in Drosophila. Cell 145(3):371–382. https://doi.org/10.1016/j.cell.2011.03.021

Baek GH, Kim I, Rao H (2011) The Cdc48 ATPase modulates the interaction between two proteolytic factors Ufd2 and Rad23. Proc Natl Acad Sci U S A 108(33):13558–13563. https://doi.org/10.1073/pnas.1104051108

Baek K, Krist DT, Prabu JR, Hill S, Klugel M, Neumaier LM, von Gronau S, Kleiger G, Schulman BA (2020) NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly. Nature 578(7795):461–466. https://doi.org/10.1038/s41586-020-2000-y

Bajorek M, Finley D, Glickman MH (2003) Proteasome disassembly and downregulation is correlated with viability during stationary phase. Curr Biol 13(13):1140–1144. https://doi.org/10.1016/s0960-9822(03)00417-2

Baldwin AS (2001) Control of oncogenesis and cancer therapy resistance by the transcription factor NF-kappaB. J Clin Invest 107(3):241–246. https://doi.org/10.1172/JCI11991

Banerjee S, Bartesaghi A, Merk A, Rao P, Bulfer SL, Yan Y, Green N, Mroczkowski B, Neitz RJ, Wipf P, Falconieri V, Deshaies RJ, Milne JL, Huryn D, Arkin M, Subramaniam S (2016) 2.3 A resolution cryo-EM structure of human p 97 and mechanism of allosteric inhibition. Science 351(6275):871–875. https://doi.org/10.1126/science.aad7974

Baraibar MA, Liu L, Ahmed EK, Friguet B (2012) Protein oxidative damage at the crossroads of cellular senescence, aging, and age-related diseases. Oxid Med Cell Longev 2012:919832. https://doi.org/10.1155/2012/919832

Bard JAM, Bashore C, Dong KC, Martin A (2019) The 26S Proteasome utilizes a kinetic gateway to prioritize substrate degradation. Cell177(2):286–298 e215. https://doi.org/10.1016/j.cell.2019.02.031

Bard JAM, Goodall EA, Greene ER, Jonsson E, Dong KC, Martin A (2018) Structure and function of the 26S proteasome. Annu Rev Biochem 87:697–724. https://doi.org/10.1146/annurev-biochem-062917-011931

Barthelme D, Chen JZ, Grabenstatter J, Baker TA, Sauer RT (2014) Architecture and assembly of the archaeal Cdc48*20S proteasome. Proc Natl Acad Sci U S A 111(17):E1687–E1694. https://doi.org/10.1073/pnas.1404823111

Barthelme D, Sauer RT (2012) Identification of the Cdc48*20S proteasome as an ancient AAA + proteolytic machine. Science 337(6096):843–846. https://doi.org/10.1126/science.1224352

Barthelme D, Sauer RT (2013) Bipartite determinants mediate an evolutionarily conserved interaction between Cdc48 and the 20S peptidase. Proc Natl Acad Sci U S A 110(9):3327–3332. https://doi.org/10.1073/pnas.1300408110

Bashore C, Dambacher CM, Goodall EA, Matyskiela ME, Lander GC, Martin A (2015) Ubp6 deubiquitinase controls conformational dynamics and substrate degradation of the 26S proteasome. Nat Struct Mol Biol 22(9):712–719. https://doi.org/10.1038/nsmb.3075

Basler M, Dajee M, Moll C, Groettrup M, Kirk CJ (2010) Prevention of experimental colitis by a selective inhibitor of the immunoproteasome. J Immunol 185(1):634–641. https://doi.org/10.4049/jimmunol.0903182

Bays NW, Wilhovsky SK, Goradia A, Hodgkiss-Harlow K, Hampton RY (2001) HRD4/NPL4 is required for the proteasomal processing of ubiquitinated ER proteins. Mol Biol Cell 12(12):4114–4128. https://doi.org/10.1091/mbc.12.12.4114

Beal R, Deveraux Q, Xia G, Rechsteiner M, Pickart C (1996) Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. Proc Natl Acad Sci U S A 93(2):861–866. https://doi.org/10.1073/pnas.93.2.861

Beck F, Unverdorben P, Bohn S, Schweitzer A, Pfeifer G, Sakata E, Nickell S, Plitzko JM, Villa E, Baumeister W, Forster F (2012) Near-atomic resolution structural model of the yeast 26S proteasome. Proc Natl Acad Sci U S A 109(37):14870–14875. https://doi.org/10.1073/pnas.1213333109

Beck M, Hurt E (2017) The nuclear pore complex: understanding its function through structural insight. Nat Rev Mol Cell Biol 18(2):73–89. https://doi.org/10.1038/nrm.2016.147

Beckwith R, Estrin E, Worden EJ, Martin A (2013) Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA + unfoldase. Nat Struct Mol Biol 20(10):1164–1172. https://doi.org/10.1038/nsmb.2659

Bedford L, Hay D, Devoy A, Paine S, Powe DG, Seth R, Gray T, Topham I, Fone K, Rezvani N, Mee M, Soane T, Layfield R, Sheppard PW, Ebendal T, Usoskin D, Lowe J, Mayer RJ (2008) Depletion of 26S proteasomes in mouse brain neurons causes neurodegeneration and Lewy-like inclusions resembling human pale bodies. J Neurosci 28(33):8189–8198. https://doi.org/10.1523/JNEUROSCI.2218-08.2008

Benaroudj N, Zwickl P, Seemuller E, Baumeister W, Goldberg AL (2003) ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation. Mol Cell 11(1):69–78

Benvegnu S, Mateo MI, Palomer E, Jurado-Arjona J, Dotti CG (2017) Aging triggers cytoplasmic depletion and nuclear translocation of the E3 ligase mahogunin: A function for ubiquitin in neuronal survival. Mol Cell 66(3):358–372 e357. https://doi.org/10.1016/j.molcel.2017.04.005

Besche HC, Haas W, Gygi SP, Goldberg AL (2009) Isolation of mammalian 26S proteasomes and p97/VCP complexes using the ubiquitin-like domain from HHR23B reveals novel proteasome-associated proteins. Biochemistry 48(11):2538–2549. https://doi.org/10.1021/bi802198q

Beskow A, Grimberg KB, Bott LC, Salomons FA, Dantuma NP, Young P (2009) A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation. J Mol Biol 394(4):732–746. https://doi.org/10.1016/j.jmb.2009.09.050

Biggins S, Ivanovska I, Rose MD (1996) Yeast ubiquitin-like genes are involved in duplication of the microtubule organizing center. J Cell Biol 133(6):1331–1346. https://doi.org/10.1083/jcb.133.6.1331

Bingol B, Schuman EM (2006) Activity-dependent dynamics and sequestration of proteasomes in dendritic spines. Nature 441(7097):1144–1148. https://doi.org/10.1038/nature04769

Blanco B, Sanchez-Abarca LI, Caballero-Velazquez T, Santamaria C, Inoges S, Perez-Simon JA (2011) Depletion of alloreactive T-cells in vitro using the proteasome inhibitor bortezomib preserves the immune response against pathogens. Leuk Res 35(10):1412–1415. https://doi.org/10.1016/j.leukres.2011.05.018

Blickwedehl J, Agarwal M, Seong C, Pandita RK, Melendy T, Sung P, Pandita TK, Bangia N (2008) Role for proteasome activator PA200 and postglutamyl proteasome activity in genomic stability. Proc Natl Acad Sci U S A 105(42):16165–16170. https://doi.org/10.1073/pnas.0803145105

Bochtler M, Ditzel L, Groll M, Huber R (1997) Crystal structure of heat shock locus V (HslV) from Escherichia coli. Proc Natl Acad Sci US A 94(12):6070–6074. https://doi.org/10.1073/pnas.94.12.6070

Bodnar NO, Kim KH, Ji Z, Wales TE, Svetlov V, Nudler E, Engen JR, Walz T, Rapoport TA (2018) Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1-Npl4. Nat Struct Mol Biol 25(7):616–622. https://doi.org/10.1038/s41594-018-0085-x

Bodnar NO, Rapoport TA (2017) Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex. Cell 169(4):722–735 e729. https://doi.org/10.1016/j.cell.2017.04.020

Boehringer J, Riedinger C, Paraskevopoulos K, Johnson EO, Lowe ED, Khoudian C, Smith D, Noble ME, Gordon C, Endicott JA (2012) Structural and functional characterization of Rpn12 identifies residues required for Rpn10 proteasome incorporation. Biochem J 448(1):55–65. https://doi.org/10.1042/BJ20120542

Bogyo M, Gaczynska M, Ploegh HL (1997a) Proteasome inhibitors and antigen presentation. Biopolymers 43(4):269–280. https://doi.org/10.1002/(SICI)1097-0282(1997)43:4%3c269:AID-BIP2%3e3.0.CO;2-T

Bogyo M, McMaster JS, Gaczynska M, Tortorella D, Goldberg AL, Ploegh H (1997b) Covalent modification of the active site threonine of proteasomal beta subunits and the Escherichia coli homolog HslV by a new class of inhibitors. Proc Natl Acad Sci U S A 94(13):6629–6634. https://doi.org/10.1073/pnas.94.13.6629

Bogyo M, Shin S, McMaster JS, Ploegh HL (1998) Substrate binding and sequence preference of the proteasome revealed by active-site-directed affinity probes. Chem Biol 5(6):307–320. https://doi.org/10.1016/s1074-5521(98)90169-7

Bohn S, Beck F, Sakata E, Walzthoeni T, Beck M, Aebersold R, Forster F, Baumeister W, Nickell S (2010) Structure of the 26S proteasome from Schizosaccharomyces pombe at subnanometer resolution. Proc Natl Acad Sci U S A 107(49):20992–20997. https://doi.org/10.1073/pnas.1015530107

Boname JM, Thomas M, Stagg HR, Xu P, Peng J, Lehner PJ (2010) Efficient internalization of MHC I requires lysine-11 and lysine-63 mixed linkage polyubiquitin chains. Traffic 11(2):210–220. https://doi.org/10.1111/j.1600-0854.2009.01011.x

Book AJ, Gladman NP, Lee SS, Scalf M, Smith LM, Vierstra RD (2010) Affinity purification of the Arabidopsis 26 S proteasome reveals a diverse array of plant proteolytic complexes. J Biol Chem 285(33):25554–25569. https://doi.org/10.1074/jbc.M110.136622

Borissenko L, Groll M (2007) 20S proteasome and its inhibitors: crystallographic knowledge for drug development. Chem Rev 107(3):687–717. https://doi.org/10.1021/cr0502504

Borodovsky A, Kessler BM, Casagrande R, Overkleeft HS, Wilkinson KD, Ploegh HL (2001) A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. EMBO J 20(18):5187–5196. https://doi.org/10.1093/emboj/20.18.5187

Brandman O, Stewart-Ornstein J, Wong D, Larson A, Williams CC, Li GW, Zhou S, King D, Shen PS, Weibezahn J, Dunn JG, Rouskin S, Inada T, Frost A, Weissman JS (2012) A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell 151(5):1042–1054. https://doi.org/10.1016/j.cell.2012.10.044

Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, Murzin AG (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. Nature 378(6555):416–419. https://doi.org/10.1038/378416a0

Braten O, Livneh I, Ziv T, Admon A, Kehat I, Caspi LH, Gonen H, Bercovich B, Godzik A, Jahandideh S, Jaroszewski L, Sommer T, Kwon YT, Guharoy M, Tompa P, Ciechanover A (2016) Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination. Proc Natl Acad Sci U S A 113(32):E4639–E4647. https://doi.org/10.1073/pnas.1608644113

Braun S, Matuschewski K, Rape M, Thoms S, Jentsch S (2002) Role of the ubiquitin-selective CDC48(UFD1/NPL4)chaperone (segregase) in ERAD of OLE1 and other substrates. EMBO J 21(4):615–621. https://doi.org/10.1093/emboj/21.4.615

Bremm A, Moniz S, Mader J, Rocha S, Komander D (2014) Cezanne (OTUD7B) regulates HIF-1alpha homeostasis in a proteasome-independent manner. EMBO Rep 15(12):1268–1277. https://doi.org/10.15252/embr.201438850

Brettschneider J, Del Tredici K, Lee VM, Trojanowski JQ (2015) Spreading of pathology in neurodegenerative diseases: a focus on human studies. Nat Rev Neurosci 16(2):109–120. https://doi.org/10.1038/nrn3887

Breusing N, Grune T (2008) Regulation of proteasome-mediated protein degradation during oxidative stress and aging. Biol Chem 389(3):203–209. https://doi.org/10.1515/BC.2008.029
Brooks P, Fuertes G, Murray RZ, Bose S, Knecht E, Rechsteiner MC, Hendil KB, Tanaka K, Dyson J, Rivett J (2000) Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. Biochem J 346(Pt 1):155–161
Broz P, Dixit VM (2016) Inflammasomes: mechanism of assembly, regulation and signalling. Nat Rev Immunol 16(7):407–420. https://doi.org/10.1038/nri.2016.58
Buchberger A, Schindelin H, Hanzelmann P (2015) Control of p 97 function by cofactor binding. FEBS Lett 589(19 Pt A):2578–2589. https://doi.org/10.1016/j.febslet.2015.08.028
Budenholzer L, Cheng CL, Li Y, Hochstrasser M (2017) Proteasome structure and assembly. J Mol Biol 429(22):3500–3524. https://doi.org/10.1016/j.jmb.2017.05.027
Buel GR, Chen X, Chari R, O'Neill MJ, Ebelle DL, Jenkins C, Sridharan V, Tarasov SG, Tarasova NI, Andresson T, Walters KJ (2020) Structure of E3 ligase E6AP with a proteasome-binding site provided by substrate receptor hRpn10. Nat Commun 11(1):1291. https://doi.org/10.1038/s41467-020-15073-7
Buetow L, Huang DT (2016) Structural insights into the catalysis and regulation of E3 ubiquitin ligases. Nat Rev Mol Cell Biol 17(10):626–642. https://doi.org/10.1038/nrm.2016.91
Bug M, Meyer H (2012) Expanding into new markets–VCP/p97 in endocytosis and autophagy. J Struct Biol 179(2):78–82. https://doi.org/10.1016/j.jsb.2012.03.003
Bulteau AL, Lundberg KC, Humphries KM, Sadek HA, Szweda PA, Friguet B, Szweda LI (2001) Oxidative modification and inactivation of the proteasome during coronary occlusion/reperfusion. J Biol Chem 276(32):30057–30063. https://doi.org/10.1074/jbc.M100142200
Burcoglu J, Zhao L, Enenkel C (2015) Nuclear import of yeast proteasomes. Cells 4(3):387–405. https://doi.org/10.3390/cells4030387
Burgie SE, Bingman CA, Soni AB, Phillips GN Jr (2012) Structural characterization of human Uch37. Proteins 80(2):649–654. https://doi.org/10.1002/prot.23147
Cai ZP, Shen Z, Van Kaer L, Becker LC (2008) Ischemic preconditioning-induced cardioprotection is lost in mice with immunoproteasome subunit low molecular mass polypeptide-2 deficiency. FASEB J 22(12):4248–4257. https://doi.org/10.1096/fj.08-105940
Carvalho P, Goder V, Rapoport TA (2006) Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. Cell 126(2):361–373. https://doi.org/10.1016/j.cell.2006.05.043
Cascio P, Call M, Petre BM, Walz T, Goldberg AL (2002) Properties of the hybrid form of the 26S proteasome containing both 19S and PA28 complexes. EMBO J 21(11):2636–2645. https://doi.org/10.1093/emboj/21.11.2636
Castaneda CA, Kashyap TR, Nakasone MA, Krueger S, Fushman D (2013) Unique structural, dynamical, and functional properties of k11-linked polyubiquitin chains. Structure 21(7):1168–1181. https://doi.org/10.1016/j.str.2013.04.029
Chang CW, Lee S, Tsai FTF (2017) Structural elements regulating AAA + protein quality control machines. Front Mol Biosci 4:27. https://doi.org/10.3389/fmolb.2017.00027
Chapman E, Fry AN, Kang M (2011) The complexities of p97 function in health and disease. Mol BioSyst 7(3):700–710. https://doi.org/10.1039/c0mb00176g
Chatterjee-Kishore M, Wright KL, Ting JP, Stark GR (2000) How Stat1 mediates constitutive gene expression: a complex of unphosphorylated Stat1 and IRF1 supports transcription of the LMP2 gene. EMBO J 19(15):4111–4122. https://doi.org/10.1093/emboj/19.15.4111
Chen C, Edelstein LC, Gelinas C (2000) The Rel/NF-kappaB family directly activates expression of the apoptosis inhibitor Bcl-x(L). Mol Cell Biol 20(8):2687–2695. https://doi.org/10.1128/mcb.20.8.2687-2695.2000
Chen L, Madura K (2002) Rad23 promotes the targeting of proteolytic substrates to the proteasome. Mol Cell Biol 22(13):4902–4913. https://doi.org/10.1128/mcb.22.13.4902-4913.2002
Chen L, Madura K (2005) Increased proteasome activity, ubiquitin-conjugating enzymes, and eEF1A translation factor detected in breast cancer tissue. Cancer Res 65(13):5599–5606. https://doi.org/10.1158/0008-5472.CAN-05-0201

Chen L, Madura K (2014) Yeast importin-alpha (Srp1) performs distinct roles in the import of nuclear proteins and in targeting proteasomes to the nucleus. J Biol Chem 289(46):32339–32352. https://doi.org/10.1074/jbc.M114.582023

Chen L, Romero L, Chuang SM, Tournier V, Joshi KK, Lee JA, Kovvali G, Madura K (2011) Sts1 plays a key role in targeting proteasomes to the nucleus. J Biol Chem 286(4):3104–3118. https://doi.org/10.1074/jbc.M110.135863

Chen P, Hochstrasser M (1995) Biogenesis, structure and function of the yeast 20S proteasome. EMBO J 14(11):2620–2630

Chen P, Hochstrasser M (1996) Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. Cell 86(6):961–972. https://doi.org/10.1016/s0092-8674(00)80171-3

Chen Q, Thorpe J, Dohmen JR, Li F, Keller JN (2006) Ump1 extends yeast lifespan and enhances viability during oxidative stress: central role for the proteasome? Free Radic Biol Med 40(1):120–126. https://doi.org/10.1016/j.freeradbiomed.2005.08.048

Chen S, Wu J, Lu Y, Ma YB, Lee BH, Yu Z, Ouyang Q, Finley DJ, Kirschner MW, Mao Y (2016a) Structural basis for dynamic regulation of the human 26S proteasome. Proc Natl Acad Sci U S A 113(46):12991–12996. https://doi.org/10.1073/pnas.1614614113

Chen X, Barton LF, Chi Y, Clurman BE, Roberts JM (2007) Ubiquitin-independent degradation of cell-cycle inhibitors by the REGgamma proteasome. Mol Cell 26(6):843–852. https://doi.org/10.1016/j.molcel.2007.05.022

Chen X, Ebelle DL, Wright BJ, Sridharan V, Hooper E, Walters KJ (2019) Structure of hRpn10 bound to UBQLN2 UBL illustrates basis for complementarity between shuttle factors and substrates at the proteasome. J Mol Biol 431(5):939–955. https://doi.org/10.1016/j.jmb.2019.01.021

Chen X, Lee BH, Finley D, Walters KJ (2010) Structure of proteasome ubiquitin receptor hRpn13 and its activation by the scaffolding protein hRpn2. Mol Cell 38(3):404–415. https://doi.org/10.1016/j.molcel.2010.04.019

Chen X, Randles L, Shi K, Tarasov SG, Aihara H, Walters KJ (2016b) Structures of Rpn1 T1:Rad23 and hRpn13:hPLIC2 reveal distinct binding mechanisms between substrate receptors and shuttle factors of the proteasome. Structure 24(8):1257–1270. https://doi.org/10.1016/j.str.2016.05.018

Chen Z, Hagler J, Palombella VJ, Melandri F, Scherer D, Ballard D, Maniatis T (1995) Signal-induced site-specific phosphorylation targets I kappa B alpha to the ubiquitin-proteasome pathway. Genes Dev 9(13):1586–1597. https://doi.org/10.1101/gad.9.13.1586

Chen ZJ, Sun LJ (2009) Nonproteolytic functions of ubiquitin in cell signaling. Mol Cell 33(3):275–286. https://doi.org/10.1016/j.molcel.2009.01.014

Chen X, Dorris Z, Shi D, Huang RK, Khant H, Fox T, de Val N, Williams D, Zhang P, Walters KJ (2020) Cryo-EM Reveals Unanchored M1-Ubiquitin Chain Binding at hRpn11 of the 26S Proteasome. Structure. https://doi.org/10.1016/j.str.2020.07.011

Chondrogianni N, Petropoulos I, Franceschi C, Friguet B, Gonos ES (2000) Fibroblast cultures from healthy centenarians have an active proteasome. Exp Gerontol 35(6–7):721–728. https://doi.org/10.1016/s0531-5565(00)00137-6

Christianson JC, Ye Y (2014) Cleaning up in the endoplasmic reticulum: ubiquitin in charge. Nat Struct Mol Biol 21(4):325–335. https://doi.org/10.1038/nsmb.2793

Chui AJ, Okondo MC, Rao SD, Gai K, Griswold AR, Johnson DC, Ball DP, Taabazuing CY, Orth EL, Vittimberga BA, Bachovchin DA (2019) N-terminal degradation activates the NLRP1B inflammasome. Science 364(6435):82–85. https://doi.org/10.1126/science.aau1208

Ciechanover A (2005) Proteolysis: from the lysosome to ubiquitin and the proteasome. Nat Rev Mol Cell Biol 6(1):79–87. https://doi.org/10.1038/nrm1552

Ciechanover A, Brundin P (2003) The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. Neuron 40(2):427–446. https://doi.org/10.1016/s0896-6273(03)00606-8

Ciechanover A, Elias S, Heller H, Ferber S, Hershko A (1980a) Characterization of the heat-stable polypeptide of the ATP-dependent proteolytic system from reticulocytes. J Biol Chem 255(16):7525–7528

Ciechanover A, Heller H, Elias S, Haas AL, Hershko A (1980b) ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. Proc Natl Acad Sci U S A 77(3):1365–1368. https://doi.org/10.1073/pnas.77.3.1365

Ciechanover A, Kwon YT (2015) Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. Exp Mol Med 47:e147. https://doi.org/10.1038/emm.2014.117

Ciehanover A, Hod Y, Hershko A (1978) A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. Biochem Biophys Res Commun 81(4):1100–1105. https://doi.org/10.1016/0006-291x(78)91249-4

Clague MJ, Urbe S, Komander D (2019) Breaking the chains: deubiquitylating enzyme specificity begets function. Nat Rev Mol Cell Biol 20(6):338–352. https://doi.org/10.1038/s41580-019-0099-1

Cohen-Kaplan V, Livneh I, Avni N, Fabre B, Ziv T, Kwon YT, Ciechanover A (2016) p62- and ubiquitin-dependent stress-induced autophagy of the mammalian 26S proteasome. Proc Natl Acad Sci U S A 113(47):E7490–E7499. https://doi.org/10.1073/pnas.1615455113

Collins GA, Goldberg AL (2017) The logic of the 26S proteasome. Cell 169(5):792–806. https://doi.org/10.1016/j.cell.2017.04.023

Collins GA, Goldberg AL (2020) Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation. Proc Natl Acad Sci U S A 117(9):4664–4674. https://doi.org/10.1073/pnas.1915534117

Cook WJ, Jeffrey LC, Carson M, Chen Z, Pickart CM (1992) Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). J Biol Chem 267(23):16467–16471. https://doi.org/10.2210/pdb1aar/pdb

Cooney I, Han H, Stewart MG, Carson RH, Hansen DT, Iwasa JH, Price JC, Hill CP, Shen PS (2019) Structure of the Cdc48 segregase in the act of unfolding an authentic substrate. Science 365(6452):502–505. https://doi.org/10.1126/science.aax0486

Cope GA, Suh GS, Aravind L, Schwarz SE, Zipursky SL, Koonin EV, Deshaies RJ (2002) Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1. Science 298(5593):608–611. https://doi.org/10.1126/science.1075901

Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 65:801–847. https://doi.org/10.1146/annurev.bi.65.070196.004101

Craney A, Rape M (2013) Dynamic regulation of ubiquitin-dependent cell cycle control. Curr Opin Cell Biol 25(6):704–710. https://doi.org/10.1016/j.ceb.2013.07.004

Crosas B, Hanna J, Kirkpatrick DS, Zhang DP, Tone Y, Hathaway NA, Buecker C, Leggett DS, Schmidt M, King RW, Gygi SP, Finley D (2006) Ubiquitin chains are remodeled at the proteasome by opposing ubiquitin ligase and deubiquitinating activities. Cell 127(7):1401–1413. https://doi.org/10.1016/j.cell.2006.09.051

Cruz M, Elenich LA, Smolarek TA, Menon AG, Monaco JJ (1997) DNA sequence, chromosomal localization, and tissue expression of the mouse proteasome subunit lmp10 (Psmb10) gene. Genomics 45(3):618–622. https://doi.org/10.1006/geno.1997.4977

Cui Z, Scruggs SB, Gilda JE, Ping P, Gomes AV (2014) Regulation of cardiac proteasomes by ubiquitination, SUMOylation, and beyond. J Mol Cell Cardiol 71:32–42. https://doi.org/10.1016/j.yjmcc.2013.10.008

da Fonseca PC, He J, Morris EP (2012) Molecular model of the human 26S proteasome. Mol Cell 46(1):54–66. https://doi.org/10.1016/j.molcel.2012.03.026

Dahlmann B, Ruppert T, Kuehn L, Merforth S, Kloetzel PM (2000) Different proteasome subtypes in a single tissue exhibit different enzymatic properties. J Mol Biol 303(5):643–653. https://doi.org/10.1006/jmbi.2000.4185

Dambacher CM, Worden EJ, Herzik MA, Martin A, Lander GC (2016) Atomic structure of the 26S proteasome lid reveals the mechanism of deubiquitinase inhibition. Elife 5:e13027. https://doi.org/10.7554/eLife.13027

Dammer EB, Na CH, Xu P, Seyfried NT, Duong DM, Cheng D, Gearing M, Rees H, Lah JJ, Levey AI, Rush J, Peng J (2011) Polyubiquitin linkage profiles in three models of proteolytic stress

suggest the etiology of Alzheimer disease. J Biol Chem 286(12):10457–10465. https://doi.org/10.1074/jbc.M110.149633
Dang FW, Chen L, Madura K (2016) Catalytically active proteasomes function predominantly in the cytosol. J Biol Chem 291(36):18765–18777. https://doi.org/10.1074/jbc.M115.712406
Dange T, Smith D, Noy T, Rommel PC, Jurzitza L, Cordero RJ, Legendre A, Finley D, Goldberg AL, Schmidt M (2011) Blm10 protein promotes proteasomal substrate turnover by an active gating mechanism. J Biol Chem 286(50):42830–42839. https://doi.org/10.1074/jbc.M111.300178
Dantuma NP, Hoppe T (2012) Growing sphere of influence: Cdc48/p97 orchestrates ubiquitin-dependent extraction from chromatin. Trends Cell Biol 22(9):483–491. https://doi.org/10.1016/j.tcb.2012.06.003
Darwin KH (2009) Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. Nat Rev Microbiol 7(7):485–491. https://doi.org/10.1038/nrmicro2148
Datta AB, Hura GL, Wolberger C (2009) The structure and conformation of Lys63-linked tetraubiquitin. J Mol Biol 392(5):1117–1124. https://doi.org/10.1016/j.jmb.2009.07.090
Davies CW, Paul LN, Kim MI, Das C (2011) Structural and thermodynamic comparison of the catalytic domain of AMSH and AMSH-LP: nearly identical fold but different stability. J Mol Biol 413(2):416–429. https://doi.org/10.1016/j.jmb.2011.08.029
Davies JM, Brunger AT, Weis WI (2008) Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. Structure 16(5):715–726. https://doi.org/10.1016/j.str.2008.02.010
Davies KJ (2001) Degradation of oxidized proteins by the 20S proteasome. Biochimie 83(3–4):301–310. https://doi.org/10.1016/s0300-9084(01)01250-0
De La Mota-Peynado A, Lee SY, Pierce BM, Wani P, Singh CR, Roelofs J (2013) The proteasome-associated protein Ecm29 inhibits proteasomal ATPase activity and in vivo protein degradation by the proteasome. J Biol Chem 288(41):29467–29481. https://doi.org/10.1074/jbc.M113.491662
de la Pena AH, Goodall EA, Gates SN, Lander GC, Martin A (2018) Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. Science 362(6418). https://doi.org/10.1126/science.aav0725
de Poot SAH, Tian G, Finley D (2017) Meddling with fate: the proteasomal deubiquitinating enzymes. J Mol Biol 429(22):3525–3545. https://doi.org/10.1016/j.jmb.2017.09.015
Delic J, Masdehors P, Omura S, Cosset JM, Dumont J, Binet JL, Magdelenat H (1998) The proteasome inhibitor lactacystin induces apoptosis and sensitizes chemo- and radioresistant human chronic lymphocytic leukaemia lymphocytes to TNF-alpha-initiated apoptosis. Br J Cancer 77(7):1103–1107. https://doi.org/10.1038/bjc.1998.183
Demartino GN, Gillette TG (2007) Proteasomes: machines for all reasons. Cell 129(4):659–662. https://doi.org/10.1016/j.cell.2007.05.007
Demasi M, Silva GM, Netto LE (2003) 20 S proteasome from Saccharomyces cerevisiae is responsive to redox modifications and is S-glutathionylated. J Biol Chem 278(1):679–685. https://doi.org/10.1074/jbc.M209282200
des Georges A, Dhote V, Kuhn L, Hellen CU, Pestova TV, Frank J, Hashem Y (2015) Structure of mammalian eIF3 in the context of the 43S preinitiation complex. Nature 525(7570):491–495. https://doi.org/10.1038/nature14891
Deveraux Q, Ustrell V, Pickart C, Rechsteiner M (1994) A 26 S protease subunit that binds ubiquitin conjugates. J Biol Chem 269(10):7059–7061
Deville C, Carroni M, Franke KB, Topf M, Bukau B, Mogk A, Saibil HR (2017) Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase. Sci Adv 3(8):e1701726. https://doi.org/10.1126/sciadv.1701726
Deville C, Franke K, Mogk A, Bukau B, Saibil HR (2019) Two-step activation mechanism of the ClpB disaggregase for sequential substrate threading by the main ATPase motor. Cell Rep 27(12):3433–3446 e3434. https://doi.org/10.1016/j.celrep.2019.05.075
Di Fiore PP, Polo S, Hofmann K (2003) When ubiquitin meets ubiquitin receptors: a signalling connection. Nat Rev Mol Cell Biol 4(6):491–497. https://doi.org/10.1038/nrm1124

Dibble CC, Manning BD (2013) Signal integration by mTORC1 coordinates nutrient input with biosynthetic output. Nat Cell Biol 15(6):555–564. https://doi.org/10.1038/ncb2763

Dick LR, Cruikshank AA, Grenier L, Melandri FD, Nunes SL, Stein RL (1996) Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for clasto-lactacystin beta-lactone. J Biol Chem 271(13):7273–7276. https://doi.org/10.1074/jbc.271.13.7273

Dick LR, Moomaw CR, DeMartino GN, Slaughter CA (1991) Degradation of oxidized insulin B chain by the multiproteinase complex macropain (proteasome). Biochemistry 30(10):2725–2734. https://doi.org/10.1021/bi00224a022

Dick TP, Nussbaum AK, Deeg M, Heinemeyer W, Groll M, Schirle M, Keilholz W, Stevanovic S, Wolf DH, Huber R, Rammensee HG, Schild H (1998) Contribution of proteasomal beta-subunits to the cleavage of peptide substrates analyzed with yeast mutants. J Biol Chem 273(40):25637–25646. https://doi.org/10.1074/jbc.273.40.25637

Dietrich C, Bartsch T, Schanz F, Oesch F, Wieser RJ (1996) p53-dependent cell cycle arrest induced by N-acetyl-L-leucinyl-L-leucinyl-L-norleucinal in platelet-derived growth factor-stimulated human fibroblasts. Proc Natl Acad Sci U S A 93(20):10815–10819. https://doi.org/10.1073/pnas.93.20.10815

Ding Z, Fu Z, Xu C, Wang Y, Wang Y, Li J, Kong L, Chen J, Li N, Zhang R, Cong Y (2017) High-resolution cryo-EM structure of the proteasome in complex with ADP-AlFx. Cell Res 27(3):373–385. https://doi.org/10.1038/cr.2017.12

Ding Z, Xu C, Sahu I, Wang Y, Fu Z, Huang M, Wong CCL, Glickman MH, Cong Y (2019) Structural snapshots of 26S proteasome reveal tetraubiquitin-induced conformations. Mol Cell 73(6):1150–1161 e1156. https://doi.org/10.1016/j.molcel.2019.01.018

Ditzel L, Huber R, Mann K, Heinemeyer W, Wolf DH, Groll M (1998) Conformational constraints for protein self-cleavage in the proteasome. J Mol Biol 279(5):1187–1191. https://doi.org/10.1006/jmbi.1998.1818

Djakovic SN, Marquez-Lona EM, Jakawich SK, Wright R, Chu C, Sutton MA, Patrick GN (2012) Phosphorylation of Rpt6 regulates synaptic strength in hippocampal neurons. J Neurosci 32(15):5126–5131. https://doi.org/10.1523/JNEUROSCI.4427-11.2012

Dodson G, Wlodawer A (1998) Catalytic triads and their relatives. Trends Biochem Sci 23(9):347–352. https://doi.org/10.1016/s0968-0004(98)01254-7

Dolenc I, Seemuller E, Baumeister W (1998) Decelerated degradation of short peptides by the 20S proteasome. FEBS Lett 434(3):357–361. https://doi.org/10.1016/s0014-5793(98)01010-2

Dong Y, Zhang S, Wu Z, Li X, Wang WL, Zhu Y, Stoilova-McPhie S, Lu Y, Finley D, Mao Y (2019) Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. Nature 565(7737):49–55. https://doi.org/10.1038/s41586-018-0736-4

Drews O, Taegtmeyer H (2014) Targeting the ubiquitin-proteasome system in heart disease: the basis for new therapeutic strategies. Antioxid Redox Signal 21(17):2322–2343. https://doi.org/10.1089/ars.2013.5823

Dubiel W, Pratt G, Ferrell K, Rechsteiner M (1992) Purification of an 11 S regulator of the multicatalytic protease. J Biol Chem 267(31):22369–22377

Duennwald ML, Lindquist S (2008) Impaired ERAD and ER stress are early and specific events in polyglutamine toxicity. Genes Dev 22(23):3308–3319. https://doi.org/10.1101/gad.1673408

Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC (1995) Penicillin acylase has a single-amino-acid catalytic centre. Nature 373(6511):264–268. https://doi.org/10.1038/373264a0

Dulic V, Kaufmann WK, Wilson SJ, Tlsty TD, Lees E, Harper JW, Elledge SJ, Reed SI (1994) p53-dependent inhibition of cyclin-dependent kinase activities in human fibroblasts during radiation-induced G1 arrest. Cell 76(6):1013–1023. https://doi.org/10.1016/0092-8674(94)90379-4

Dynek JN, Goncharov T, Dueber EC, Fedorova AV, Izrael-Tomasevic A, Phu L, Helgason E, Fairbrother WJ, Deshayes K, Kirkpatrick DS, Vucic D (2010) c-IAP1 and UbcH5 promote K11-linked polyubiquitination of RIP1 in TNF signalling. EMBO J 29(24):4198–4209. https://doi.org/10.1038/emboj.2010.300

Edbauer D, Haass C (2016) An amyloid-like cascade hypothesis for C9orf72 ALS/FTD. Curr Opin Neurobiol 36:99–106. https://doi.org/10.1016/j.conb.2015.10.009

Eddins MJ, Varadan R, Fushman D, Pickart CM, Wolberger C (2007) Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. J Mol Biol 367(1):204–211. https://doi.org/10.1016/j.jmb.2006.12.065

Ehring B, Meyer TH, Eckerskorn C, Lottspeich F, Tampe R (1996) Effects of major-histocompatibility-complex-encoded subunits on the peptidase and proteolytic activities of human 20S proteasomes. Cleavage of proteins and antigenic peptides. Eur J Biochem 235(1–2):404–415. https://doi.org/10.1111/j.1432-1033.1996.00404.x

Eisele MR, Reed RG, Rudack T, Schweitzer A, Beck F, Nagy I, Pfeifer G, Plitzko JM, Baumeister W, Tomko RJ, Jr., Sakata E (2018) Expanded Coverage of the 26S Proteasome conformational landscape reveals mechanisms of peptidase gating. Cell Rep 24(5):1301–1315 e1305. https://doi.org/10.1016/j.celrep.2018.07.004

Elsasser S, Chandler-Militello D, Muller B, Hanna J, Finley D (2004) Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome. J Biol Chem 279(26):26817–26822. https://doi.org/10.1074/jbc.M404020200

Elsasser S, Finley D (2005) Delivery of ubiquitinated substrates to protein-unfolding machines. Nat Cell Biol 7(8):742–749. https://doi.org/10.1038/ncb0805-742

Elsasser S, Gali RR, Schwickart M, Larsen CN, Leggett DS, Muller B, Feng MT, Tubing F, Dittmar GA, Finley D (2002) Proteasome subunit Rpn1 binds ubiquitin-like protein domains. Nat Cell Biol 4(9):725–730. https://doi.org/10.1038/ncb845

Enenkel C, Blobel G, Rexach M (1995) Identification of a yeast karyopherin heterodimer that targets import substrate to mammalian nuclear pore complexes. J Biol Chem 270(28):16499–16502. https://doi.org/10.1074/jbc.270.28.16499

Enenkel C, Lehmann A, Kloetzel PM (1998) Subcellular distribution of proteasomes implicates a major location of protein degradation in the nuclear envelope-ER network in yeast. EMBO J 17(21):6144–6154. https://doi.org/10.1093/emboj/17.21.6144

Enenkel C, Lehmann H, Kipper J, Guckel R, Hilt W, Wolf DH (1994) PRE3, highly homologous to the human major histocompatibility complex-linked LMP2 (RING12) gene, codes for a yeast proteasome subunit necessary for the peptidylglutamyl-peptide hydrolyzing activity. FEBS Lett 341(2–3):193–196. https://doi.org/10.1016/0014-5793(94)80455-9

Erales J, Coffino P (2014) Ubiquitin-independent proteasomal degradation. Biochim Biophys Acta 1843(1):216–221. https://doi.org/10.1016/j.bbamcr.2013.05.008

Erales J, Hoyt MA, Troll F, Coffino P (2012) Functional asymmetries of proteasome translocase pore. J Biol Chem 287(22):18535–18543. https://doi.org/10.1074/jbc.M112.357327

Erpapazoglou Z, Walker O, Haguenauer-Tsapis R (2014) Versatile roles of k63-linked ubiquitin chains in trafficking. Cells 3(4):1027–1088. https://doi.org/10.3390/cells3041027

Erzberger JP, Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA + proteins. Annu Rev Biophys Biomol Struct 35:93–114. https://doi.org/10.1146/annurev.biophys.35.040405.101933

Estrin E, Lopez-Blanco JR, Chacon P, Martin A (2013) Formation of an intricate helical bundle dictates the assembly of the 26S proteasome lid. Structure 21(9):1624–1635. https://doi.org/10.1016/j.str.2013.06.023

Etlinger JD, Goldberg AL (1977) A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. Proc Natl Acad Sci U S A 74(1):54–58. https://doi.org/10.1073/pnas.74.1.54

Farre JC, Subramani S (2016) Mechanistic insights into selective autophagy pathways: lessons from yeast. Nat Rev Mol Cell Biol 17(9):537–552. https://doi.org/10.1038/nrm.2016.74

Fei X, Bell TA, Jenni S, Stinson BM, Baker TA, Harrison SC, Sauer RT (2020) Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. Elife 9. 10.7554/eLife.52774

Fenteany G, Schreiber SL (1998) Lactacystin, proteasome function, and cell fate. J Biol Chem 273(15):8545–8548. https://doi.org/10.1074/jbc.273.15.8545

Fenteany G, Standaert RF, Lane WS, Choi S, Corey EJ, Schreiber SL (1995) Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. Science 268(5211):726–731. https://doi.org/10.1126/science.7732382

Ferrington DA, Gregerson DS (2012) Immunoproteasomes: structure, function, and antigen presentation. Prog Mol Biol Transl Sci 109:75–112. https://doi.org/10.1016/B978-0-12-397863-9.00003-1

Figueiredo-Pereira ME, Berg KA, Wilk S (1994) A new inhibitor of the chymotrypsin-like activity of the multicatalytic proteinase complex (20S proteasome) induces accumulation of ubiquitin-protein conjugates in a neuronal cell. J Neurochem 63(4):1578–1581. https://doi.org/10.1046/j.1471-4159.1994.63041578.x

Finley D (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annu Rev Biochem 78:477–513. https://doi.org/10.1146/annurev.biochem.78.081507.101607

Finley D, Chen X, Walters KJ (2016) Gates, channels, and switches: elements of the proteasome machine. Trends Biochem Sci 41(1):77–93. https://doi.org/10.1016/j.tibs.2015.10.009

Finley D, Prado MA (2019) The proteasome and its network: engineering for adaptability. Cold Spring Harb Perspect Biol. https://doi.org/10.1101/cshperspect.a033985

Fischer DF, van Dijk R, van Tijn P, Hobo B, Verhage MC, van der Schors RC, Li KW, van Minnen J, Hol EM, van Leeuwen FW (2009) Long-term proteasome dysfunction in the mouse brain by expression of aberrant ubiquitin. Neurobiol Aging 30(6):847–863. https://doi.org/10.1016/j.neurobiolaging.2008.06.009

Fishbain S, Inobe T, Israeli E, Chavali S, Yu H, Kago G, Babu MM, Matouschek A (2015) Sequence composition of disordered regions fine-tunes protein half-life. Nat Struct Mol Biol 22(3):214–221. https://doi.org/10.1038/nsmb.2958

Fishbain S, Prakash S, Herrig A, Elsasser S, Matouschek A (2011) Rad23 escapes degradation because it lacks a proteasome initiation region. Nat Commun 2:192. https://doi.org/10.1038/ncomms1194

Fletcher BR, Hill GS, Long JM, Gallagher M, Shapiro ML, Rapp PR (2014) A fine balance: regulation of hippocampal Arc/Arg3.1 transcription, translation and degradation in a rat model of normal cognitive aging. Neurobiol Learn Mem 115:58–67. https://doi.org/10.1016/j.nlm.2014.08.007

Flick K, Ouni I, Wohlschlegel JA, Capati C, McDonald WH, Yates JR, Kaiser P (2004) Proteolysis-independent regulation of the transcription factor Met4 by a single Lys 48-linked ubiquitin chain. Nat Cell Biol 6(7):634–641. https://doi.org/10.1038/ncb1143

Fonseca R, Vabulas RM, Hartl FU, Bonhoeffer T, Nagerl UV (2006) A balance of protein synthesis and proteasome-dependent degradation determines the maintenance of LTP. Neuron 52(2):239–245. https://doi.org/10.1016/j.neuron.2006.08.015

Foresti O, Rodriguez-Vaello V, Funaya C, Carvalho P (2014) Quality control of inner nuclear membrane proteins by the Asi complex. Science 346(6210):751–755. https://doi.org/10.1126/science.1255638

Forster A, Masters EI, Whitby FG, Robinson H, Hill CP (2005) The 1.9 A structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. Mol Cell 18(5):589–599. https://doi.org/10.1016/j.molcel.2005.04.016

Forster A, Whitby FG, Hill CP (2003) The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation. EMBO J 22(17):4356–4364. https://doi.org/10.1093/emboj/cdg436

Forster F, Unverdorben P, Sledz P, Baumeister W (2013) Unveiling the long-held secrets of the 26S proteasome. Structure 21(9):1551–1562. https://doi.org/10.1016/j.str.2013.08.010

Frank J (2006) Three-dimensional electron microscopy of macromolecular assemblies : visualization of biological molecules in their native state, 2nd edn. Oxford University Press, Oxford, New York

Frankland-Searby S, Bhaumik SR (2012) The 26S proteasome complex: an attractive target for cancer therapy. Biochim Biophys Acta 1825(1):64–76. 10.1016/j.bbcan.2011.10.003

Franz A, Ackermann L, Hoppe T (2016) Ring of change: CDC48/p97 drives protein dynamics at chromatin. Front Genet 7:73. https://doi.org/10.3389/fgene.2016.00073

Freibaum BD, Taylor JP (2017) The role of dipeptide repeats in C9ORF72-related ALS-FTD. Front Mol Neurosci 10:35. https://doi.org/10.3389/fnmol.2017.00035

Frentzel S, Pesold-Hurt B, Seelig A, Kloetzel PM (1994) 20 S proteasomes are assembled via distinct precursor complexes. Processing of LMP2 and LMP7 proproteins takes place in 13–16 S preproteasome complexes. J Mol Biol 236(4):975–981. https://doi.org/10.1016/0022-2836(94)90003-5

Fu H, Reis N, Lee Y, Glickman MH, Vierstra RD (2001) Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome. EMBO J 20(24):7096–7107. https://doi.org/10.1093/emboj/20.24.7096

Fu H, Sadis S, Rubin DM, Glickman M, van Nocker S, Finley D, Vierstra RD (1998) Multiubiquitin chain binding and protein degradation are mediated by distinct domains within the 26 S proteasome subunit Mcb1. J Biol Chem 273(4):1970–1981. https://doi.org/10.1074/jbc.273.4.1970

Fu X, Sokolova V, Webb KJ, Old W, Park S (2018) Ubiquitin-dependent switch during assembly of the proteasomal ATPases mediated by Not4 ubiquitin ligase. Proc Natl Acad Sci U S A 115(52):13246–13251. https://doi.org/10.1073/pnas.1805353115

Fukunaga K, Kudo T, Toh-e A, Tanaka K, Saeki Y (2010) Dissection of the assembly pathway of the proteasome lid in Saccharomyces cerevisiae. Biochem Biophys Res Commun 396(4):1048–1053. https://doi.org/10.1016/j.bbrc.2010.05.061

Funakoshi M, Sasaki T, Nishimoto T, Kobayashi H (2002) Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. Proc Natl Acad Sci U S A 99(2):745–750. https://doi.org/10.1073/pnas.012585199

Gan J, Leestemaker Y, Sapmaz A, Ovaa H (2019) Highlighting the proteasome: using fluorescence to visualize proteasome activity and distribution. Front Mol Biosci 6:14. https://doi.org/10.3389/fmolb.2019.00014

Gao Y, Cui Y, Fox T, Lin S, Wang H, de Val N, Zhou ZH, Yang W (2019) Structures and operating principles of the replisome. Science 363 (6429). https://doi.org/10.1126/science.aav7003

Gardner RG, Nelson ZW, Gottschling DE (2005) Degradation-mediated protein quality control in the nucleus. Cell 120(6):803–815. https://doi.org/10.1016/j.cell.2005.01.016

Gates SN, Yokom AL, Lin J, Jackrel ME, Rizo AN, Kendsersky NM, Buell CE, Sweeny EA, Mack KL, Chuang E, Torrente MP, Su M, Shorter J, Southworth DR (2017) Ratchet-like polypeptide translocation mechanism of the AAA + disaggregase Hsp104. Science 357(6348):273–279. https://doi.org/10.1126/science.aan1052

Gatica D, Lahiri V, Klionsky DJ (2018) Cargo recognition and degradation by selective autophagy. Nat Cell Biol 20(3):233–242. https://doi.org/10.1038/s41556-018-0037-z

Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, Daughrity L, Dunmore JH, Castanedes-Casey M, Chew J, Cosio DM, van Blitterswijk M, Lee WC, Rademakers R, Boylan KB, Dickson DW, Petrucelli L (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol 126(6):829–844. https://doi.org/10.1007/s00401-013-1192-8

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425(6959):737–741. https://doi.org/10.1038/nature02046

Gladman NP, Marshall RS, Lee KH, Vierstra RD (2016) The proteasome stress regulon is controlled by a pair of NAC transcription factors in Arabidopsis. Plant Cell 28(6):1279–1296. https://doi.org/10.1105/tpc.15.01022

Glickman MH, Rubin DM, Coux O, Wefes I, Pfeifer G, Cjeka Z, Baumeister W, Fried VA, Finley D (1998) A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. Cell 94(5):615–623. https://doi.org/10.1016/s0092-8674(00)81603-7

Glynn SE, Martin A, Nager AR, Baker TA, Sauer RT (2009) Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA + protein-unfolding machine. Cell 139(4):744–756. https://doi.org/10.1016/j.cell.2009.09.034

Godderz D, Heinen C, Marchese FP, Kurz T, Acs K, Dantuma NP (2015) Cdc48-independent proteasomal degradation coincides with a reduced need for ubiquitylation. Sci Rep 5:7615. https://doi.org/10.1038/srep07615

Goebl MG, Goetsch L, Byers B (1994) The Ubc3 (Cdc34) ubiquitin-conjugating enzyme is ubiquitinated and phosphorylated in vivo. Mol Cell Biol 14(5):3022–3029. https://doi.org/10.1128/mcb.14.5.3022

Goldstein G, Scheid M, Hammerling U, Schlesinger DH, Niall HD, Boyse EA (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. Proc Natl Acad Sci U S A 72(1):11–15. https://doi.org/10.1073/pnas.72.1.11

Gomes AV, Zong C, Edmondson RD, Li X, Stefani E, Zhang J, Jones RC, Thyparambil S, Wang GW, Qiao X, Bardag-Gorce F, Ping P (2006) Mapping the murine cardiac 26S proteasome complexes. Circ Res 99(4):362–371. https://doi.org/10.1161/01.RES.0000237386.98506.f7

Gomez TA, Kolawa N, Gee M, Sweredoski MJ, Deshaies RJ (2011) Identification of a functional docking site in the Rpn1 LRR domain for the UBA-UBL domain protein Ddi1. BMC Biol 9:33. https://doi.org/10.1186/1741-7007-9-33

Gorbea C, Goellner GM, Teter K, Holmes RK, Rechsteiner M (2004) Characterization of mammalian Ecm29, a 26 S proteasome-associated protein that localizes to the nucleus and membrane vesicles. J Biol Chem 279(52):54849–54861. https://doi.org/10.1074/jbc.M410444200

Gorbea C, Pratt G, Ustrell V, Bell R, Sahasrabudhe S, Hughes RE, Rechsteiner M (2010) A protein interaction network for Ecm29 links the 26 S proteasome to molecular motors and endosomal components. J Biol Chem 285(41):31616–31633. https://doi.org/10.1074/jbc.M110.154120

Greene ER, Goodall EA, de la Pena AH, Matyskiela ME, Lander GC, Martin A (2019) Specific lid-base contacts in the 26 s proteasome control the conformational switching required for substrate degradation. Elife 8. https://doi.org/10.7554/elife.49806

Groll M, Bajorek M, Kohler A, Moroder L, Rubin DM, Huber R, Glickman MH, Finley D (2000) A gated channel into the proteasome core particle. Nat Struct Biol 7(11):1062–1067. https://doi.org/10.1038/80992

Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. Nature 386(6624):463–471. https://doi.org/10.1038/386463a0

Groll M, Heinemeyer W, Jager S, Ullrich T, Bochtler M, Wolf DH, Huber R (1999) The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. Proc Natl Acad Sci U S A 96(20):10976–10983. https://doi.org/10.1073/pnas.96.20.10976

Groll M, Huber R (2004) Inhibitors of the eukaryotic 20S proteasome core particle: a structural approach. Biochim Biophys Acta 1695(1–3):33–44. https://doi.org/10.1016/j.bbamcr.2004.09.025

Grune T, Jung T, Merker K, Davies KJ (2004) Decreased proteolysis caused by protein aggregates, inclusion bodies, plaques, lipofuscin, ceroid, and 'aggresomes' during oxidative stress, aging, and disease. Int J Biochem Cell Biol 36(12):2519–2530. https://doi.org/10.1016/j.biocel.2004.04.020

Gu ZC, Wu E, Sailer C, Jando J, Styles E, Eisenkolb I, Kuschel M, Bitschar K, Wang X, Huang L, Vissa A, Yip CM, Yedidi RS, Friesen H, Enenkel C (2017) Ubiquitin orchestrates proteasome dynamics between proliferation and quiescence in yeast. Mol Biol Cell 28(19):2479–2491. https://doi.org/10.1091/mbc.E17-03-0162

Guerrero C, Tagwerker C, Kaiser P, Huang L (2006) An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. Mol Cell Proteomics 5(2):366–378. https://doi.org/10.1074/mcp.M500303-MCP200

Guo Q, Lehmer C, Martinez-Sanchez A, Rudack T, Beck F, Hartmann H, Perez-Berlanga M, Frottin F, Hipp MS, Hartl FU, Edbauer D, Baumeister W, Fernandez-Busnadiego R (2018) In situ structure of neuronal C9orf72 Poly-GA aggregates reveals proteasome recruitment. Cell 172(4):696–705 e612. https://doi.org/10.1016/j.cell.2017.12.030

Guo X, Engel JL, Xiao J, Tagliabracci VS, Wang X, Huang L, Dixon JE (2011) UBLCP1 is a 26S proteasome phosphatase that regulates nuclear proteasome activity. Proc Natl Acad Sci U S A 108(46):18649–18654. https://doi.org/10.1073/pnas.1113170108

Guterman A, Glickman MH (2004) Complementary roles for Rpn11 and Ubp6 in deubiquitination and proteolysis by the proteasome. J Biol Chem 279(3):1729–1738. https://doi.org/10.1074/jbc.M307050200

Haapasalo A, Viswanathan J, Kurkinen KM, Bertram L, Soininen H, Dantuma NP, Tanzi RE, Hiltunen M (2011) Involvement of ubiquilin-1 transcript variants in protein degradation and accumulation. Commun Integr Biol 4(4):428–432. https://doi.org/10.4161/cib.4.4.15283

Hahn JS, Neef DW, Thiele DJ (2006) A stress regulatory network for co-ordinated activation of proteasome expression mediated by yeast heat shock transcription factor. Mol Microbiol 60(1):240–251. https://doi.org/10.1111/j.1365-2958.2006.05097.x

Hallermalm K, Seki K, Wei C, Castelli C, Rivoltini L, Kiessling R, Levitskaya J (2001) Tumor necrosis factor-alpha induces coordinated changes in major histocompatibility class I presentation pathway, resulting in increased stability of class I complexes at the cell surface. Blood 98(4):1108–1115. https://doi.org/10.1182/blood.v98.4.1108

Hamazaki J, Hirayama S, Murata S (2015) Redundant roles of Rpn10 and Rpn13 in recognition of ubiquitinated proteins and cellular homeostasis. PLoS Genet 11(7):e1005401. https://doi.org/10.1371/journal.pgen.1005401

Hamazaki J, Iemura S, Natsume T, Yashiroda H, Tanaka K, Murata S (2006) A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. EMBO J 25(19):4524–4536. https://doi.org/10.1038/sj.emboj.7601338

Han H, Fulcher JM, Dandey VP, Iwasa JH, Sundquist WI, Kay MS, Shen PS, Hill CP (2019) Structure of Vps4 with circular peptides and implications for translocation of two polypeptide chains by AAA + ATPases. Elife 8. https://doi.org/10.7554/elife.44071

Han H, Monroe N, Sundquist WI, Shen PS, Hill CP (2017) The AAA ATPase Vps4 binds ESCRT-III substrates through a repeating array of dipeptide-binding pockets. Elife 6. https://doi.org/10.7554/elife.31324

Hanna J, Hathaway NA, Tone Y, Crosas B, Elsasser S, Kirkpatrick DS, Leggett DS, Gygi SP, King RW, Finley D (2006) Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. Cell 127(1):99–111. https://doi.org/10.1016/j.cell.2006.07.038

Hanson PI, Whiteheart SW (2005) AAA + proteins: have engine, will work. Nat Rev Mol Cell Biol 6(7):519–529. https://doi.org/10.1038/nrm1684

Hanzelmann P, Schindelin H (2017) The interplay of cofactor interactions and post-translational modifications in the regulation of the AAA + ATPase p97. Front Mol Biosci 4:21. https://doi.org/10.3389/fmolb.2017.00021

Hanzelmann P, Stingele J, Hofmann K, Schindelin H, Raasi S (2010) The yeast E4 ubiquitin ligase Ufd2 interacts with the ubiquitin-like domains of Rad23 and Dsk2 via a novel and distinct ubiquitin-like binding domain. J Biol Chem 285(26):20390–20398. https://doi.org/10.1074/jbc.M110.112532

Haratake K, Sato A, Tsuruta F, Chiba T (2016) KIAA0368-deficiency affects disassembly of 26S proteasome under oxidative stress condition. J Biochem 159(6):609–618. https://doi.org/10.1093/jb/mvw006

Harding CV, France J, Song R, Farah JM, Chatterjee S, Iqbal M, Siman R (1995) Novel dipeptide aldehydes are proteasome inhibitors and block the MHC-I antigen-processing pathway. J Immunol 155(4):1767–1775

Harris JR (1968) Release of a macromolecular protein component from human erythrocyte ghosts. Biochim Biophys Acta 150(3):534–537. https://doi.org/10.1016/0005-2736(68)90157-0

Harshbarger W, Miller C, Diedrich C, Sacchettini J (2015) Crystal structure of the human 20S proteasome in complex with carfilzomib. Structure 23(2):418–424. https://doi.org/10.1016/j.str.2014.11.017

Haselbach D, Schrader J, Lambrecht F, Henneberg F, Chari A, Stark H (2017) Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. Nat Commun 8:15578. https://doi.org/10.1038/ncomms15578

Hayashi M, Ishibashi T, Tanaka K, Kasahara M (1997) The mouse genes encoding the third pair of beta-type proteasome subunits regulated reciprocally by IFN-gamma: structural comparison, chromosomal localization, and analysis of the promoter. J Immunol 159(6):2760–2770

He J, Kulkarni K, da Fonseca PC, Krutauz D, Glickman MH, Barford D, Morris EP (2012) The structure of the 26S proteasome subunit Rpn2 reveals its PC repeat domain as a closed toroid of two concentric alpha-helical rings. Structure 20(3):513–521. https://doi.org/10.1016/j.str.2011.12.015

Hegde AN, Smith SG, Duke LM, Pourquoi A, Vaz S (2019) Perturbations of ubiquitin-proteasome-mediated proteolysis in aging and alzheimer's disease. Front Aging Neurosci 11:324. https://doi.org/10.3389/fnagi.2019.00324

Heinemeyer W, Fischer M, Krimmer T, Stachon U, Wolf DH (1997) The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. J Biol Chem 272(40):25200–25209. https://doi.org/10.1074/jbc.272.40.25200

Heinemeyer W, Gruhler A, Mohrle V, Mahe Y, Wolf DH (1993) PRE2, highly homologous to the human major histocompatibility complex-linked RING10 gene, codes for a yeast proteasome subunit necessary for chrymotryptic activity and degradation of ubiquitinated proteins. J Biol Chem 268(7):5115–5120

Heinemeyer W, Kleinschmidt JA, Saidowsky J, Escher C, Wolf DH (1991) Proteinase yscE, the yeast proteasome/multicatalytic-multifunctional proteinase: mutants unravel its function in stress induced proteolysis and uncover its necessity for cell survival. EMBO J 10(3):555–562

Heinen C, Acs K, Hoogstraten D, Dantuma NP (2011) C-terminal UBA domains protect ubiquitin receptors by preventing initiation of protein degradation. Nat Commun 2:191. https://doi.org/10.1038/ncomms1179

Heo JM, Livnat-Levanon N, Taylor EB, Jones KT, Dephoure N, Ring J, Xie J, Brodsky JL, Madeo F, Gygi SP, Ashrafi K, Glickman MH, Rutter J (2010) A stress-responsive system for mitochondrial protein degradation. Mol Cell 40(3):465–480. https://doi.org/10.1016/j.molcel.2010.10.021

Hershko A, Ciechanover A, Heller H, Haas AL, Rose IA (1980) Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. Proc Natl Acad Sci U S A 77(4):1783–1786. https://doi.org/10.1073/pnas.77.4.1783

Hershko A, Ciechanover A, Rose IA (1979) Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP. Proc Natl Acad Sci U S A 76(7):3107–3110. https://doi.org/10.1073/pnas.76.7.3107

Hilt W, Enenkel C, Gruhler A, Singer T, Wolf DH (1993) The PRE4 gene codes for a subunit of the yeast proteasome necessary for peptidylglutamyl-peptide-hydrolyzing activity. Mutations link the proteasome to stress- and ubiquitin-dependent proteolysis. J Biol Chem 268(5):3479–3486

Hinnerwisch J, Fenton WA, Furtak KJ, Farr GW, Horwich AL (2005) Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. Cell 121(7):1029–1041. https://doi.org/10.1016/j.cell.2005.04.012

Hirano H, Kimura Y, Kimura A (2016) Biological significance of co- and post-translational modifications of the yeast 26S proteasome. J Proteomics 134:37–46. https://doi.org/10.1016/j.jprot.2015.11.016

Hirano Y, Hayashi H, Iemura S, Hendil KB, Niwa S, Kishimoto T, Kasahara M, Natsume T, Tanaka K, Murata S (2006) Cooperation of multiple chaperones required for the assembly of mammalian 20S proteasomes. Mol Cell 24(6):977–984. https://doi.org/10.1016/j.molcel.2006.11.015

Hirano Y, Hendil KB, Yashiroda H, Iemura S, Nagane R, Hioki Y, Natsume T, Tanaka K, Murata S (2005) A heterodimeric complex that promotes the assembly of mammalian 20S proteasomes. Nature 437(7063):1381–1385. https://doi.org/10.1038/nature04106

Hirano Y, Kaneko T, Okamoto K, Bai M, Yashiroda H, Furuyama K, Kato K, Tanaka K, Murata S (2008) Dissecting beta-ring assembly pathway of the mammalian 20S proteasome. EMBO J 27(16):2204–2213. https://doi.org/10.1038/emboj.2008.148

Hisamatsu H, Shimbara N, Saito Y, Kristensen P, Hendil KB, Fujiwara T, Takahashi E, Tanahashi N, Tamura T, Ichihara A, Tanaka K (1996) Newly identified pair of proteasomal subunits regulated reciprocally by interferon gamma. J Exp Med 183(4):1807–1816. https://doi.org/10.1084/jem.183.4.1807

Hishiya A, Iemura S, Natsume T, Takayama S, Ikeda K, Watanabe K (2006) A novel ubiquitin-binding protein ZNF216 functioning in muscle atrophy. EMBO J 25(3):554–564. https://doi.org/10.1038/sj.emboj.7600945

Hjerpe R, Bett JS, Keuss MJ, Solovyova A, McWilliams TG, Johnson C, Sahu I, Varghese J, Wood N, Wightman M, Osborne G, Bates GP, Glickman MH, Trost M, Knebel A, Marchesi F, Kurz T (2016) UBQLN2 mediates autophagy-independent protein aggregate clearance by the proteasome. Cell 166(4):935–949. https://doi.org/10.1016/j.cell.2016.07.001

Hnia K, Clausen T, Moog-Lutz C (2019) Shaping striated muscles with ubiquitin proteasome system in health and disease. Trends Mol Med. https://doi.org/10.1016/j.molmed.2019.05.008

Hoeller D, Dikic I (2009) Targeting the ubiquitin system in cancer therapy. Nature 458(7237):438–444. https://doi.org/10.1038/nature07960

Hoffman L, Rechsteiner M (1996) Nucleotidase activities of the 26 S proteasome and its regulatory complex. J Biol Chem 271(51):32538–32545

Hofmann K, Bucher P (1996) The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. Trends Biochem Sci 21(5):172–173

Hofmann K, Bucher P (1998) The PCI domain: a common theme in three multiprotein complexes. Trends Biochem Sci 23(6):204–205. https://doi.org/10.1016/s0968-0004(98)01217-1

Hofmann RM, Pickart CM (2001) In vitro assembly and recognition of Lys-63 polyubiquitin chains. J Biol Chem 276(30):27936–27943. https://doi.org/10.1074/jbc.M103378200

Holzl H, Kapelari B, Kellermann J, Seemuller E, Sumegi M, Udvardy A, Medalia O, Sperling J, Muller SA, Engel A, Baumeister W (2000) The regulatory complex of Drosophila melanogaster 26S proteasomes. Subunit composition and localization of a deubiquitylating enzyme. J Cell Biol 150(1):119–130. https://doi.org/10.1083/jcb.150.1.119

Hoppe T, Matuschewski K, Rape M, Schlenker S, Ulrich HD, Jentsch S (2000) Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. Cell 102(5):577–586. https://doi.org/10.1016/s0092-8674(00)00080-5

Hoppe T, Rape M, Jentsch S (2001) Membrane-bound transcription factors: regulated release by RIP or RUP. Curr Opin Cell Biol 13(3):344–348. https://doi.org/10.1016/s0955-0674(00)00218-0

Hosokawa N, Hara T, Kaizuka T, Kishi C, Takamura A, Miura Y, Iemura S, Natsume T, Takehana K, Yamada N, Guan JL, Oshiro N, Mizushima N (2009) Nutrient-dependent mTORC1 association with the ULK1-Atg13-FIP200 complex required for autophagy. Mol Biol Cell 20(7):1981–1991. https://doi.org/10.1091/mbc.E08-12-1248

Hospenthal MK, Freund SM, Komander D (2013) Assembly, analysis and architecture of atypical ubiquitin chains. Nat Struct Mol Biol 20(5):555–565. https://doi.org/10.1038/nsmb.2547

Hough R, Pratt G, Rechsteiner M (1986) Ubiquitin-lysozyme conjugates. Identification and characterization of an ATP-dependent protease from rabbit reticulocyte lysates. J Biol Chem 261(5):2400–2408

Hough R, Pratt G, Rechsteiner M (1987) Purification of two high molecular weight proteases from rabbit reticulocyte lysate. J Biol Chem 262(17):8303–8313

Hoyt MA, Zich J, Takeuchi J, Zhang M, Govaerts C, Coffino P (2006) Glycine-alanine repeats impair proper substrate unfolding by the proteasome. EMBO J 25(8):1720–1729. https://doi.org/10.1038/sj.emboj.7601058

Hua Z, Vierstra RD (2011) The cullin-RING ubiquitin-protein ligases. Annu Rev Plant Biol 62:299–334. https://doi.org/10.1146/annurev-arplant-042809-112256

Huang X, Luan B, Wu J, Shi Y (2016) An atomic structure of the human 26S proteasome. Nat Struct Mol Biol 23(9):778–785. https://doi.org/10.1038/nsmb.3273

Huber EM, Basler M, Schwab R, Heinemeyer W, Kirk CJ, Groettrup M, Groll M (2012) Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. Cell 148(4):727–738. https://doi.org/10.1016/j.cell.2011.12.030

Huber EM, Groll M (2012) Inhibitors for the immuno- and constitutive proteasome: current and future trends in drug development. Angew Chem Int Ed Engl 51(35):8708–8720. https://doi.org/10.1002/anie.201201616

Huber EM, Heinemeyer W, Li X, Arendt CS, Hochstrasser M, Groll M (2016) A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. Nat Commun 7:10900. https://doi.org/10.1038/ncomms10900

Hurley JH, Stenmark H (2011) Molecular mechanisms of ubiquitin-dependent membrane traffic. Annu Rev Biophys 40:119–142. https://doi.org/10.1146/annurev-biophys-042910-155404

Husnjak K, Dikic I (2012) Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. Annu Rev Biochem 81:291–322. https://doi.org/10.1146/annurev-biochem-051810-094654

Husnjak K, Elsasser S, Zhang N, Chen X, Randles L, Shi Y, Hofmann K, Walters KJ, Finley D, Dikic I (2008) Proteasome subunit Rpn13 is a novel ubiquitin receptor. Nature 453(7194):481–488. https://doi.org/10.1038/nature06926

Hussong SA, Kapphahn RJ, Phillips SL, Maldonado M, Ferrington DA (2010) Immunoproteasome deficiency alters retinal proteasome's response to stress. J Neurochem 113(6):1481–1490. https://doi.org/10.1111/j.1471-4159.2010.06688.x

Huyer G, Piluek WF, Fansler Z, Kreft SG, Hochstrasser M, Brodsky JL, Michaelis S (2004) Distinct machinery is required in Saccharomyces cerevisiae for the endoplasmic reticulum-associated degradation of a multispanning membrane protein and a soluble luminal protein. J Biol Chem 279(37):38369–38378. https://doi.org/10.1074/jbc.M402468200

Inn KS, Gack MU, Tokunaga F, Shi M, Wong LY, Iwai K, Jung JU (2011) Linear ubiquitin assembly complex negatively regulates RIG-I- and TRIM25-mediated type I interferon induction. Mol Cell 41(3):354–365. https://doi.org/10.1016/j.molcel.2010.12.029

Inobe T, Fishbain S, Prakash S, Matouschek A (2011) Defining the geometry of the two-component proteasome degron. Nat Chem Biol 7(3):161–167. https://doi.org/10.1038/nchembio.521

Iosefson O, Nager AR, Baker TA, Sauer RT (2015) Coordinated gripping of substrate by subunits of a AAA + proteolytic machine. Nat Chem Biol 11(3):201–206. https://doi.org/10.1038/nchembio.1732

Isasa M, Katz EJ, Kim W, Yugo V, Gonzalez S, Kirkpatrick DS, Thomson TM, Finley D, Gygi SP, Crosas B (2010) Monoubiquitination of RPN10 regulates substrate recruitment to the proteasome. Mol Cell 38(5):733–745. https://doi.org/10.1016/j.molcel.2010.05.001

Ishii T, Funakoshi M, Kobayashi H (2006) Yeast Pth2 is a UBL domain-binding protein that participates in the ubiquitin-proteasome pathway. EMBO J 25(23):5492–5503. https://doi.org/10.1038/sj.emboj.7601418

Isono E, Nishihara K, Saeki Y, Yashiroda H, Kamata N, Ge L, Ueda T, Kikuchi Y, Tanaka K, Nakano A, Toh-e A (2007) The assembly pathway of the 19S regulatory particle of the yeast 26S proteasome. Mol Biol Cell 18(2):569–580. https://doi.org/10.1091/mbc.e06-07-0635

Itakura E, Zavodszky E, Shao S, Wohlever ML, Keenan RJ, Hegde RS (2016) Ubiquilins chaperone and triage mitochondrial membrane proteins for degradation. Mol Cell 63(1):21–33. https://doi.org/10.1016/j.molcel.2016.05.020

Iwafune Y, Kawasaki H, Hirano H (2002) Electrophoretic analysis of phosphorylation of the yeast 20S proteasome. Electrophoresis 23(2):329–338. https://doi.org/10.1002/1522-2683(200202)23:2%3c329:AID-ELPS329%3e3.0.CO;2-1

Iwai K, Fujita H, Sasaki Y (2014) Linear ubiquitin chains: NF-kappaB signalling, cell death and beyond. Nat Rev Mol Cell Biol 15(8):503–508. https://doi.org/10.1038/nrm3836

Iwanczyk J, Sadre-Bazzaz K, Ferrell K, Kondrashkina E, Formosa T, Hill CP, Ortega J (2006) Structure of the Blm10–20 S proteasome complex by cryo-electron microscopy. Insights into the mechanism of activation of mature yeast proteasomes. J Mol Biol 363(3):648–659. https://doi.org/10.1016/j.jmb.2006.08.010

Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA + ATPases. J Struct Biol 146(1–2):11–31. https://doi.org/10.1016/j.jsb.2003.10.010

Jacobson AD, MacFadden A, Wu Z, Peng J, Liu CW (2014) Autoregulation of the 26S proteasome by in situ ubiquitination. Mol Biol Cell 25(12):1824–1835. https://doi.org/10.1091/mbc.E13-10-0585

Jagannath S, Barlogie B, Berenson J, Siegel D, Irwin D, Richardson PG, Niesvizky R, Alexanian R, Limentani SA, Alsina M, Adams J, Kauffman M, Esseltine DL, Schenkein DP, Anderson KC (2004) A phase 2 study of two doses of bortezomib in relapsed or refractory myeloma. Br J Haematol 127(2):165–172. https://doi.org/10.1111/j.1365-2141.2004.05188.x

James AB, Conway AM, Morris BJ (2006) Regulation of the neuronal proteasome by Zif268 (Egr1). J Neurosci 26(5):1624–1634. https://doi.org/10.1523/JNEUROSCI.4199-05.2006

Janse DM, Crosas B, Finley D, Church GM (2004) Localization to the proteasome is sufficient for degradation. J Biol Chem 279(20):21415–21420. https://doi.org/10.1074/jbc.M402954200

Jarosch E, Taxis C, Volkwein C, Bordallo J, Finley D, Wolf DH, Sommer T (2002) Protein dislocation from the ER requires polyubiquitination and the AAA-ATPase Cdc48. Nat Cell Biol 4(2):134–139. https://doi.org/10.1038/ncb746

Jin L, Williamson A, Banerjee S, Philipp I, Rape M (2008) Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. Cell 133(4):653–665. https://doi.org/10.1016/j.cell.2008.04.012

Ju JS, Fuentealba RA, Miller SE, Jackson E, Piwnica-Worms D, Baloh RH, Weihl CC (2009) Valosin-containing protein (VCP) is required for autophagy and is disrupted in VCP disease. J Cell Biol 187(6):875–888. https://doi.org/10.1083/jcb.200908115

Kaganovich D, Kopito R, Frydman J (2008) Misfolded proteins partition between two distinct quality control compartments. Nature 454(7208):1088–1095. https://doi.org/10.1038/nature07195

Kaiser SE, Riley BE, Shaler TA, Trevino RS, Becker CH, Schulman H, Kopito RR (2011) Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. Nat Methods 8(8):691–696. https://doi.org/10.1038/nmeth.1649

Kaneko T, Hamazaki J, Iemura S, Sasaki K, Furuyama K, Natsume T, Tanaka K, Murata S (2009) Assembly pathway of the Mammalian proteasome base subcomplex is mediated by multiple specific chaperones. Cell 137(5):914–925. https://doi.org/10.1016/j.cell.2009.05.008

Kaplun L, Tzirkin R, Bakhrat A, Shabek N, Ivantsiv Y, Raveh D (2005) The DNA damage-inducible UbL-UbA protein Ddi1 participates in Mec1-mediated degradation of Ho endonuclease. Mol Cell Biol 25(13):5355–5362. https://doi.org/10.1128/MCB.25.13.5355-5362.2005

Karmon O, Ben Aroya S (2019) Spatial organization of proteasome aggregates in the regulation of proteasome homeostasis. Front Mol Biosci 6:150. https://doi.org/10.3389/fmolb.2019.00150

Kay LE (2011) Solution NMR spectroscopy of supra-molecular systems, why bother?. A methyl-TROSY view. J Magn Reson 210(2):159–170. https://doi.org/10.1016/j.jmr.2011.03.008

Keller JN, Hanni KB, Markesbery WR (2000) Impaired proteasome function in Alzheimer's disease. J Neurochem 75(1):436–439. https://doi.org/10.1046/j.1471-4159.2000.0750436.x

Khajuria RK, Munschauer M, Ulirsch JC, Fiorini C, Ludwig LS, McFarland SK, Abdulhay NJ, Specht H, Keshishian H, Mani DR, Jovanovic M, Ellis SR, Fulco CP, Engreitz JM, Schutz S, Lian J, Gripp KW, Weinberg OK, Pinkus GS, Gehrke L, Regev A, Lander ES, Gazda HT, Lee WY, Panse VG, Carr SA, Sankaran VG (2018) Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. Cell 173(1):90–103 e119. https://doi.org/10.1016/j.cell.2018.02.036

Khare S, Nagle AS, Biggart A, Lai YH, Liang F, Davis LC, Barnes SW, Mathison CJ, Myburgh E, Gao MY, Gillespie JR, Liu X, Tan JL, Stinson M, Rivera IC, Ballard J, Yeh V, Groessl T, Federe G, Koh HX, Venable JD, Bursulaya B, Shapiro M, Mishra PK, Spraggon G, Brock A, Mottram JC, Buckner FS, Rao SP, Wen BG, Walker JR, Tuntland T, Molteni V, Glynne RJ, Supek F (2016) Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. Nature 537(7619):229–233. https://doi.org/10.1038/nature19339

Khmelinskii A, Blaszczak E, Pantazopoulou M, Fischer B, Omnus DJ, Le Dez G, Brossard A, Gunnarsson A, Barry JD, Meurer M, Kirrmaier D, Boone C, Huber W, Rabut G, Ljungdahl PO,

Knop M (2014) Protein quality control at the inner nuclear membrane. Nature 516(7531):410–413. https://doi.org/10.1038/nature14096

Kikuchi J, Iwafune Y, Akiyama T, Okayama A, Nakamura H, Arakawa N, Kimura Y, Hirano H (2010) Co- and post-translational modifications of the 26S proteasome in yeast. Proteomics 10(15):2769–2779. https://doi.org/10.1002/pmic.200900283

Kim DY, Scalf M, Smith LM, Vierstra RD (2013) Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. Plant Cell 25(5):1523–1540. https://doi.org/10.1105/tpc.112.108613

Kim HT, Goldberg AL (2017) The deubiquitinating enzyme Usp14 allosterically inhibits multiple proteasomal activities and ubiquitin-independent proteolysis. J Biol Chem 292(23):9830–9839. https://doi.org/10.1074/jbc.M116.763128

Kim HT, Goldberg AL (2018) UBL domain of Usp14 and other proteins stimulates proteasome activities and protein degradation in cells. Proc Natl Acad Sci U S A 115(50):E11642–E11650. https://doi.org/10.1073/pnas.1808731115

Kim I, Ahn J, Liu C, Tanabe K, Apodaca J, Suzuki T, Rao H (2006) The Png1-Rad23 complex regulates glycoprotein turnover. J Cell Biol 172(2):211–219. https://doi.org/10.1083/jcb.200507149

Kim I, Mi K, Rao H (2004) Multiple interactions of rad23 suggest a mechanism for ubiquitylated substrate delivery important in proteolysis. Mol Biol Cell 15(7):3357–3365. https://doi.org/10.1091/mbc.e03-11-0835

Kim S, Saeki Y, Fukunaga K, Suzuki A, Takagi K, Yamane T, Tanaka K, Mizushima T, Kato K (2010) Crystal structure of yeast rpn14, a chaperone of the 19 S regulatory particle of the proteasome. J Biol Chem 285(20):15159–15166. https://doi.org/10.1074/jbc.M110.104042

Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, Harper JW, Gygi SP (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell 44(2):325–340. https://doi.org/10.1016/j.molcel.2011.08.025

Kim YC, DeMartino GN (2011) C termini of proteasomal ATPases play nonequivalent roles in cellular assembly of mammalian 26 S proteasome. J Biol Chem 286(30):26652–26666. https://doi.org/10.1074/jbc.M111.246793

Kim YC, Snoberger A, Schupp J, Smith DM (2015) ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. Nat Commun 6:8520. https://doi.org/10.1038/ncomms9520

Kimonis VE, Fulchiero E, Vesa J, Watts G (2008) VCP disease associated with myopathy, Paget disease of bone and frontotemporal dementia: review of a unique disorder. Biochim Biophys Acta 1782(12):744–748. https://doi.org/10.1016/j.bbadis.2008.09.003

Kimura A, Kato Y, Hirano H (2012) N-myristoylation of the Rpt2 subunit regulates intracellular localization of the yeast 26S proteasome. Biochemistry 51(44):8856–8866. https://doi.org/10.1021/bi3007862

Kimura A, Kurata Y, Nakabayashi J, Kagawa H, Hirano H (2016) N-Myristoylation of the Rpt2 subunit of the yeast 26S proteasome is implicated in the subcellular compartment-specific protein quality control system. J Proteomics 130:33–41. https://doi.org/10.1016/j.jprot.2015.08.021

Kimura Y, Kurata Y, Ishikawa A, Okayama A, Kamita M, Hirano H (2013) N-Terminal methylation of proteasome subunit Rpt1 in yeast. Proteomics 13(21):3167–3174. https://doi.org/10.1002/pmic.201300207

King RW, Deshaies RJ, Peters JM, Kirschner MW (1996) How proteolysis drives the cell cycle. Science 274(5293):1652–1659. https://doi.org/10.1126/science.274.5293.1652

Kirisako T, Kamei K, Murata S, Kato M, Fukumoto H, Kanie M, Sano S, Tokunaga F, Tanaka K, Iwai K (2006) A ubiquitin ligase complex assembles linear polyubiquitin chains. EMBO J 25(20):4877–4887. https://doi.org/10.1038/sj.emboj.7601360

Kirkpatrick DS, Hathaway NA, Hanna J, Elsasser S, Rush J, Finley D, King RW, Gygi SP (2006) Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. Nat Cell Biol 8(7):700–710. https://doi.org/10.1038/ncb1436

Kisselev AF (2013) A novel bullet hits the proteasome. Cancer Cell 24(6):691–693. https://doi.org/10.1016/j.ccr.2013.11.016

Kisselev AF, Akopian TN, Castillo V, Goldberg AL (1999) Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. Mol Cell 4(3):395–402. https://doi.org/10.1016/s1097-2765(00)80341-x

Kisselev AF, Akopian TN, Goldberg AL (1998) Range of sizes of peptide products generated during degradation of different proteins by archaeal proteasomes. J Biol Chem 273(4):1982–1989. https://doi.org/10.1074/jbc.273.4.1982

Kisselev AF, Goldberg AL (2001) Proteasome inhibitors: from research tools to drug candidates. Chem Biol 8(8):739–758. https://doi.org/10.1016/s1074-5521(01)00056-4

Kleijnen MF, Roelofs J, Park S, Hathaway NA, Glickman M, King RW, Finley D (2007) Stability of the proteasome can be regulated allosterically through engagement of its proteolytic active sites. Nat Struct Mol Biol 14(12):1180–1188. https://doi.org/10.1038/nsmb1335

Klionsky DJ, Schulman BA (2014) Dynamic regulation of macroautophagy by distinctive ubiquitin-like proteins. Nat Struct Mol Biol 21(4):336–345. https://doi.org/10.1038/nsmb.2787

Kloetzel PM (2001) Antigen processing by the proteasome. Nat Rev Mol Cell Biol 2(3):179–187. https://doi.org/10.1038/35056572

Knowlton JR, Johnston SC, Whitby FG, Realini C, Zhang Z, Rechsteiner M, Hill CP (1997) Structure of the proteasome activator REGalpha (PA28alpha). Nature 390(6660):639–643. https://doi.org/10.1038/37670

Ko HS, Uehara T, Tsuruma K, Nomura Y (2004) Ubiquilin interacts with ubiquitylated proteins and proteasome through its ubiquitin-associated and ubiquitin-like domains. FEBS Lett 566(1–3):110–114. https://doi.org/10.1016/j.febslet.2004.04.031

Kock M, Nunes MM, Hemann M, Kube S, Dohmen RJ, Herzog F, Ramos PC, Wendler P (2015) Proteasome assembly from 15S precursors involves major conformational changes and recycling of the Pba1-Pba2 chaperone. Nat Commun 6:6123. https://doi.org/10.1038/ncomms7123

Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, Jentsch S (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell 96(5):635–644. https://doi.org/10.1016/s0092-8674(00)80574-7

Komander D, Clague MJ, Urbe S (2009) Breaking the chains: structure and function of the deubiquitinases. Nat Rev Mol Cell Biol 10(8):550–563. https://doi.org/10.1038/nrm2731

Komander D, Rape M (2012) The ubiquitin code. Annu Rev Biochem 81:203–229. https://doi.org/10.1146/annurev-biochem-060310-170328

Kopp F, Dahlmann B, Kuehn L (2001) Reconstitution of hybrid proteasomes from purified PA700-20 S complexes and PA28alphabeta activator: ultrastructure and peptidase activities. J Mol Biol 313(3):465–471. https://doi.org/10.1006/jmbi.2001.5063

Kors S, Geijtenbeek K, Reits E, Schipper-Krom S (2019) Regulation of proteasome activity by (Post-)transcriptional mechanisms. Front Mol Biosci 6:48. https://doi.org/10.3389/fmolb.2019.00048

Kraut DA, Israeli E, Schrader EK, Patil A, Nakai K, Nanavati D, Inobe T, Matouschek A (2012) Sequence- and species-dependence of proteasomal processivity. ACS Chem Biol 7(8):1444–1453. https://doi.org/10.1021/cb3001155

Kraut DA, Matouschek A (2011) Proteasomal degradation from internal sites favors partial proteolysis via remote domain stabilization. ACS Chem Biol 6(10):1087–1095. https://doi.org/10.1021/cb2002285

Kravtsova-Ivantsiv Y, Ciechanover A (2012) Non-canonical ubiquitin-based signals for proteasomal degradation. J Cell Sci 125(Pt 3):539–548. https://doi.org/10.1242/jcs.093567

Kravtsova-Ivantsiv Y, Cohen S, Ciechanover A (2009) Modification by single ubiquitin moieties rather than polyubiquitination is sufficient for proteasomal processing of the p105 NF-kappaB precursor. Mol Cell 33(4):496–504. https://doi.org/10.1016/j.molcel.2009.01.023

Kristariyanto YA, Abdul Rehman SA, Campbell DG, Morrice NA, Johnson C, Toth R, Kulathu Y (2015a) K29-selective ubiquitin binding domain reveals structural basis of specificity and

heterotypic nature of k29 polyubiquitin. Mol Cell 58(1):83–94. https://doi.org/10.1016/j.molcel.2015.01.041

Kristariyanto YA, Choi SY, Rehman SA, Ritorto MS, Campbell DG, Morrice NA, Toth R, Kulathu Y (2015b) Assembly and structure of Lys33-linked polyubiquitin reveals distinct conformations. Biochem J 467(2):345–352. https://doi.org/10.1042/BJ20141502

Kruegel U, Robison B, Dange T, Kahlert G, Delaney JR, Kotireddy S, Tsuchiya M, Tsuchiyama S, Murakami CJ, Schleit J, Sutphin G, Carr D, Tar K, Dittmar G, Kaeberlein M, Kennedy BK, Schmidt M (2011) Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. PLoS Genet 7(9):e1002253. https://doi.org/10.1371/journal.pgen.1002253

Krzywda S, Brzozowski AM, Higashitsuji H, Fujita J, Welchman R, Dawson S, Mayer RJ, Wilkinson AJ (2004) The crystal structure of gankyrin, an oncoprotein found in complexes with cyclin-dependent kinase 4, a 19 S proteasomal ATPase regulator, and the tumor suppressors Rb and p53. J Biol Chem 279(2):1541–1545. https://doi.org/10.1074/jbc.M310265200

Kuhnle S, Martinez-Noel G, Leclere F, Hayes SD, Harper JW, Howley PM (2018) Angelman syndrome-associated point mutations in the Zn(2 +)-binding N-terminal (AZUL) domain of UBE3A ubiquitin ligase inhibit binding to the proteasome. J Biol Chem 293(47):18387–18399. https://doi.org/10.1074/jbc.RA118.004653

Kulathu Y, Komander D (2012) Atypical ubiquitylation—the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. Nat Rev Mol Cell Biol 13(8):508–523. https://doi.org/10.1038/nrm3394

Kuo CL, Goldberg AL (2017) Ubiquitinated proteins promote the association of proteasomes with the deubiquitinating enzyme Usp14 and the ubiquitin ligase Ube3c. Proc Natl Acad Sci U S A 114(17):E3404–E3413. https://doi.org/10.1073/pnas.1701734114

Kusmierczyk AR, Hochstrasser M (2008) Some assembly required: dedicated chaperones in eukaryotic proteasome biogenesis. Biol Chem 389(9):1143–1151. https://doi.org/10.1515/BC.2008.130

Kwon YD, Nagy I, Adams PD, Baumeister W, Jap BK (2004) Crystal structures of the Rhodococcus proteasome with and without its pro-peptides: implications for the role of the pro-peptide in proteasome assembly. J Mol Biol 335(1):233–245. https://doi.org/10.1016/j.jmb.2003.08.029

Lam YA, Lawson TG, Velayutham M, Zweier JL, Pickart CM (2002) A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal. Nature 416(6882):763–767. https://doi.org/10.1038/416763a

Lam YA, Xu W, DeMartino GN, Cohen RE (1997) Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. Nature 385(6618):737–740. https://doi.org/10.1038/385737a0

Lander GC, Estrin E, Matyskiela ME, Bashore C, Nogales E, Martin A (2012) Complete subunit architecture of the proteasome regulatory particle. Nature 482(7384):186–191. https://doi.org/10.1038/nature10774

Laporte D, Salin B, Daignan-Fornier B, Sagot I (2008) Reversible cytoplasmic localization of the proteasome in quiescent yeast cells. J Cell Biol 181(5):737–745. https://doi.org/10.1083/jcb.200711154

Lasker K, Forster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A, Baumeister W (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. Proc Natl Acad Sci U S A 109(5):1380–1387. https://doi.org/10.1073/pnas.1120559109

Le Tallec B, Barrault MB, Courbeyrette R, Guerois R, Marsolier-Kergoat MC, Peyroche A (2007) 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals. Mol Cell 27(4):660–674. https://doi.org/10.1016/j.molcel.2007.06.025

Lee BH, Lee MJ, Park S, Oh DC, Elsasser S, Chen PC, Gartner C, Dimova N, Hanna J, Gygi SP, Wilson SM, King RW, Finley D (2010) Enhancement of proteasome activity by a small-molecule inhibitor of USP14. Nature 467(7312):179–184. https://doi.org/10.1038/nature09299

Lee BH, Lu Y, Prado MA, Shi Y, Tian G, Sun S, Elsasser S, Gygi SP, King RW, Finley D (2016) USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. Nature 532(7599):398–401. https://doi.org/10.1038/nature17433

Lee CK, Klopp RG, Weindruch R, Prolla TA (1999) Gene expression profile of aging and its retardation by caloric restriction. Science 285(5432):1390–1393. https://doi.org/10.1126/science.285.5432.1390

Lee D, Takayama S, Goldberg AL (2018) ZFAND5/ZNF216 is an activator of the 26S proteasome that stimulates overall protein degradation. Proc Natl Acad Sci U S A 115(41):E9550–E9559. https://doi.org/10.1073/pnas.1809934115

Lee DH, Goldberg AL (1998) Proteasome inhibitors: valuable new tools for cell biologists. Trends Cell Biol 8(10):397–403. https://doi.org/10.1016/s0962-8924(98)01346-4

Lee SY, De la Mota-Peynado A, Roelofs J (2011) Loss of Rpt5 protein interactions with the core particle and Nas2 protein causes the formation of faulty proteasomes that are inhibited by Ecm29 protein. J Biol Chem 286(42):36641–36651. https://doi.org/10.1074/jbc.M111.280875

Leggett DS, Hanna J, Borodovsky A, Crosas B, Schmidt M, Baker RT, Walz T, Ploegh H, Finley D (2002) Multiple associated proteins regulate proteasome structure and function. Mol Cell 10(3):495–507. https://doi.org/10.1016/s1097-2765(02)00638-x

Lehmann A, Janek K, Braun B, Kloetzel PM, Enenkel C (2002) 20 S proteasomes are imported as precursor complexes into the nucleus of yeast. J Mol Biol 317(3):401–413. https://doi.org/10.1006/jmbi.2002.5443

Lehmann A, Niewienda A, Jechow K, Janek K, Enenkel C (2010) Ecm29 fulfils quality control functions in proteasome assembly. Mol Cell 38(6):879–888. https://doi.org/10.1016/j.molcel.2010.06.016

Leone P, Shin EC, Perosa F, Vacca A, Dammacco F, Racanelli V (2013) MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. J Natl Cancer Inst 105(16):1172–1187. https://doi.org/10.1093/jnci/djt184

Levine B, Kroemer G (2019) Biological functions of autophagy genes: a disease perspective. Cell 176(1–2):11–42. https://doi.org/10.1016/j.cell.2018.09.048

Leznicki P, Kulathu Y (2017) Mechanisms of regulation and diversification of deubiquitylating enzyme function. J Cell Sci 130(12):1997–2006. https://doi.org/10.1242/jcs.201855

Li F, Tian G, Langager D, Sokolova V, Finley D, Park S (2017) Nucleotide-dependent switch in proteasome assembly mediated by the Nas6 chaperone. Proc Natl Acad Sci U S A 114(7):1548–1553. https://doi.org/10.1073/pnas.1612922114

Li H, O'Donoghue AJ, van der Linden WA, Xie SC, Yoo E, Foe IT, Tilley L, Craik CS, da Fonseca PC, Bogyo M (2016a) Structure- and function-based design of Plasmodium-selective proteasome inhibitors. Nature 530(7589):233–236. https://doi.org/10.1038/nature16936

Li J, Horak KM, Su H, Sanbe A, Robbins J, Wang X (2011) Enhancement of proteasomal function protects against cardiac proteinopathy and ischemia/reperfusion injury in mice. J Clin Invest 121(9):3689–3700. https://doi.org/10.1172/JCI45709

Li X, Amazit L, Long W, Lonard DM, Monaco JJ, O'Malley BW (2007a) Ubiquitin- and ATP-independent proteolytic turnover of p21 by the REGgamma-proteasome pathway. Mol Cell 26(6):831–842. https://doi.org/10.1016/j.molcel.2007.05.028

Li X, Demartino GN (2009) Variably modulated gating of the 26S proteasome by ATP and polyubiquitin. Biochem J 421(3):397–404. https://doi.org/10.1042/BJ20090528

Li X, Kusmierczyk AR, Wong P, Emili A, Hochstrasser M (2007b) beta-Subunit appendages promote 20S proteasome assembly by overcoming an Ump1-dependent checkpoint. EMBO J 26(9):2339–2349. https://doi.org/10.1038/sj.emboj.7601681

Li X, Li Y, Arendt CS, Hochstrasser M (2016b) Distinct elements in the proteasomal beta5 subunit propeptide required for autocatalytic processing and proteasome assembly. J Biol Chem 291(4):1991–2003. https://doi.org/10.1074/jbc.M115.677047

Li X, Thompson D, Kumar B, DeMartino GN (2014) Molecular and cellular roles of PI31 (PSMF1) protein in regulation of proteasome function. J Biol Chem 289(25):17392–17405. https://doi.org/10.1074/jbc.M114.561183

Li X, Wang CE, Huang S, Xu X, Li XJ, Li H, Li S (2010) Inhibiting the ubiquitin-proteasome system leads to preferential accumulation of toxic N-terminal mutant huntingtin fragments. Hum Mol Genet 19(12):2445–2455. https://doi.org/10.1093/hmg/ddq127

Li YF, Wang X (2011) The role of the proteasome in heart disease. Biochim Biophys Acta 1809(2):141–149. https://doi.org/10.1016/j.bbagrm.2010.09.001

Liang RY, Chen L, Ko BT, Shen YH, Li YT, Chen BR, Lin KT, Madura K, Chuang SM (2014) Rad23 interaction with the proteasome is regulated by phosphorylation of its ubiquitin-like (UbL) domain. J Mol Biol 426(24):4049–4060. https://doi.org/10.1016/j.jmb.2014.10.004

Lin G, Li D, de Carvalho LP, Deng H, Tao H, Vogt G, Wu K, Schneider J, Chidawanyika T, Warren JD, Li H, Nathan C (2009) Inhibitors selective for mycobacterial versus human proteasomes. Nature 461(7264):621–626. https://doi.org/10.1038/nature08357

Lin G, Mao D, Bellen HJ (2017) Amyotrophic lateral sclerosis pathogenesis converges on defects in protein homeostasis associated with TDP-43 mislocalization and proteasome-mediated degradation overload. Curr Top Dev Biol 121:111–171. https://doi.org/10.1016/bs.ctdb.2016.07.004

Lingaraju GM, Bunker RD, Cavadini S, Hess D, Hassiepen U, Renatus M, Fischer ES, Thoma NH (2014) Crystal structure of the human COP9 signalosome. Nature 512(7513):161–165. https://doi.org/10.1038/nature13566

Lipinszki Z, Kovacs L, Deak P, Udvardy A (2012) Ubiquitylation of drosophila p54/Rpn10/S5a regulates its interaction with the UBA-UBL polyubiquitin receptors. Biochemistry 51(12):2461–2470. https://doi.org/10.1021/bi3001006

Liu X, Xiao W, Zhang Y, Wiley SE, Zuo T, Zheng Y, Chen N, Chen L, Wang X, Zheng Y, Huang L, Lin S, Murphy AN, Dixon JE, Xu P, Guo X (2020) Reversible phosphorylation of Rpn1 regulates 26S proteasome assembly and function. Proc Natl Acad Sci U S A 117(1):328–336. https://doi.org/10.1073/pnas.1912531117

Liu Y, Lu L, Hettinger CL, Dong G, Zhang D, Rezvani K, Wang X, Wang H (2014) Ubiquilin-1 protects cells from oxidative stress and ischemic stroke caused tissue injury in mice. J Neurosci 34(8):2813–2821. https://doi.org/10.1523/JNEUROSCI.3541-13.2014

Liu Z, Dong X, Yi HW, Yang J, Gong Z, Wang Y, Liu K, Zhang WP, Tang C (2019) Structural basis for the recognition of K48-linked Ub chain by proteasomal receptor Rpn13. Cell Discov 5:19. https://doi.org/10.1038/s41421-019-0089-7

Liu Z, Gong Z, Jiang WX, Yang J, Zhu WK, Guo DC, Zhang WP, Liu ML, Tang C (2015) Lys63-linked ubiquitin chain adopts multiple conformational states for specific target recognition. Elife 4. https://doi.org/10.7554/elife.05767

Livneh I, Cohen-Kaplan V, Cohen-Rosenzweig C, Avni N, Ciechanover A (2016) The life cycle of the 26S proteasome: from birth, through regulation and function, and onto its death. Cell Res 26(8):869–885. https://doi.org/10.1038/cr.2016.86

Lokireddy S, Kukushkin NV, Goldberg AL (2015) cAMP-induced phosphorylation of 26S proteasomes on Rpn6/PSMD11 enhances their activity and the degradation of misfolded proteins. Proc Natl Acad Sci U S A 112(52):E7176–E7185. https://doi.org/10.1073/pnas.1522332112

Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R (1995) Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. Science 268(5210):533–539. https://doi.org/10.1126/science.7725097

Lu H, Zong C, Wang Y, Young GW, Deng N, Souda P, Li X, Whitelegge J, Drews O, Yang PY, Ping P (2008) Revealing the dynamics of the 20 S proteasome phosphoproteome: a combined CID and electron transfer dissociation approach. Mol Cell Proteomics 7(11):2073–2089. https://doi.org/10.1074/mcp.M800064-MCP200

Lu K, Psakhye I, Jentsch S (2014) Autophagic clearance of polyQ proteins mediated by ubiquitin-Atg8 adaptors of the conserved CUET protein family. Cell 158(3):549–563. https://doi.org/10.1016/j.cell.2014.05.048

Lu X, Ebelle DL, Matsuo H, Walters KJ (2020) An extended conformation for K48 ubiquitin chains revealed by the hRpn2:Rpn13:K48-Diubiquitin structure. Structure. https://doi.org/10.1016/j.str.2020.02.007

Lu X, Nowicka U, Sridharan V, Liu F, Randles L, Hymel D, Dyba M, Tarasov SG, Tarasova NI, Zhao XZ, Hamazaki J, Murata S, Burke TR Jr, Walters KJ (2017a) Structure of the Rpn13-Rpn2

complex provides insights for Rpn13 and Uch37 as anticancer targets. Nat Commun 8:15540. https://doi.org/10.1038/ncomms15540

Lu Y, Lee BH, King RW, Finley D, Kirschner MW (2015) Substrate degradation by the proteasome: a single-molecule kinetic analysis. Science 348(6231):1250834. https://doi.org/10.1126/science.1250834

Lu Y, Wu J, Dong Y, Chen S, Sun S, Ma YB, Ouyang Q, Finley D, Kirschner MW, Mao Y (2017b) Conformational landscape of the p 28-Bound human proteasome regulatory particle. Mol Cell 67(2):322–333 e326. https://doi.org/10.1016/j.molcel.2017.06.007

Luan B, Huang X, Wu J, Mei Z, Wang Y, Xue X, Yan C, Wang J, Finley DJ, Shi Y, Wang F (2016) Structure of an endogenous yeast 26S proteasome reveals two major conformational states. Proc Natl Acad Sci U S A 113(10):2642–2647. https://doi.org/10.1073/pnas.1601561113

Lupas A, Zwickl P, Wenzel T, Seemuller E, Baumeister W (1995) Structure and function of the 20S proteasome and of its regulatory complexes. Cold Spring Harb Symp Quant Biol 60:515–524. https://doi.org/10.1101/sqb.1995.060.01.055

Lupas AN, Martin J (2002) AAA proteins. Curr Opin Struct Biol 12(6):746–753

Ma CP, Slaughter CA, DeMartino GN (1992) Identification, purification, and characterization of a protein activator (PA28) of the 20 S proteasome (macropain). J Biol Chem 267(15):10515–10523

Ma M, Liu ZL (2010) Comparative transcriptome profiling analyses during the lag phase uncover YAP1, PDR1, PDR3, RPN4, and HSF1 as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for Saccharomyces cerevisiae. BMC Genom 11:660. https://doi.org/10.1186/1471-2164-11-660

Ma W, Lehner PJ, Cresswell P, Pober JS, Johnson DR (1997) Interferon-gamma rapidly increases peptide transporter (TAP) subunit expression and peptide transport capacity in endothelial cells. J Biol Chem 272(26):16585–16590. https://doi.org/10.1074/jbc.272.26.16585

Machiels BM, Henfling ME, Gerards WL, Broers JL, Bloemendal H, Ramaekers FC, Schutte B (1997) Detailed analysis of cell cycle kinetics upon proteasome inhibition. Cytometry 28(3):243–252

Magnaghi P, D'Alessio R, Valsasina B, Avanzi N, Rizzi S, Asa D, Gasparri F, Cozzi L, Cucchi U, Orrenius C, Polucci P, Ballinari D, Perrera C, Leone A, Cervi G, Casale E, Xiao Y, Wong C, Anderson DJ, Galvani A, Donati D, O'Brien T, Jackson PK, Isacchi A (2013) Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. Nat Chem Biol 9(9):548–556. https://doi.org/10.1038/nchembio.1313

Majumder P, Rudack T, Beck F, Danev R, Pfeifer G, Nagy I, Baumeister W (2019) Cryo-EM structures of the archaeal PAN-proteasome reveal an around-the-ring ATPase cycle. Proc Natl Acad Sci U S A 116(2):534–539. https://doi.org/10.1073/pnas.1817752116

Malinovska L, Kroschwald S, Munder MC, Richter D, Alberti S (2012) Molecular chaperones and stress-inducible protein-sorting factors coordinate the spatiotemporal distribution of protein aggregates. Mol Biol Cell 23(16):3041–3056. https://doi.org/10.1091/mbc.E12-03-0194

Mannhaupt G, Schnall R, Karpov V, Vetter I, Feldmann H (1999) Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. FEBS Lett 450(1–2):27–34. https://doi.org/10.1016/s0014-5793(99)00467-6

Mansour W, Nakasone MA, von Delbruck M, Yu Z, Krutauz D, Reis N, Kleifeld O, Sommer T, Fushman D, Glickman MH (2015) Disassembly of Lys11 and mixed linkage polyubiquitin conjugates provides insights into function of proteasomal deubiquitinases Rpn11 and Ubp6. J Biol Chem 290(8):4688–4704. https://doi.org/10.1074/jbc.M114.568295

Maqbool A, Hughes RK, Dagdas YF, Tregidgo N, Zess E, Belhaj K, Round A, Bozkurt TO, Kamoun S, Banfield MJ (2016) Structural basis of host autophagy-related protein 8 (ATG8) binding by the irish potato famine pathogen effector protein PexRD54. J Biol Chem 291(38):20270–20282. https://doi.org/10.1074/jbc.M116.744995

Marshall RS, Hua Z, Mali S, McLoughlin F, Vierstra RD (2019) ATG8-Binding UIM proteins define a new class of autophagy adaptors and receptors. Cell 177(3):766–781 e724. https://doi.org/10.1016/j.cell.2019.02.009

Marshall RS, Li F, Gemperline DC, Book AJ, Vierstra RD (2015) Autophagic degradation of the 26S proteasome is mediated by the dual ATG8/Ubiquitin receptor RPN10 in arabidopsis. Mol Cell 58(6):1053–1066. https://doi.org/10.1016/j.molcel.2015.04.023

Marshall RS, McLoughlin F, Vierstra RD (2016) Autophagic turnover of inactive 26S proteasomes in yeast is directed by the ubiquitin receptor Cue5 and the Hsp42 chaperone. Cell Rep 16(6):1717–1732. https://doi.org/10.1016/j.celrep.2016.07.015

Marshall RS, Vierstra RD (2015) Eat or be eaten: the autophagic plight of inactive 26S proteasomes. Autophagy 11(10):1927–1928. https://doi.org/10.1080/15548627.2015.1078961

Marshall RS, Vierstra RD (2018a) Autophagy: the master of bulk and selective recycling. Annu Rev Plant Biol 69:173–208. https://doi.org/10.1146/annurev-arplant-042817-040606

Marshall RS, Vierstra RD (2018b) Proteasome storage granules protect proteasomes from autophagic degradation upon carbon starvation. Elife 7. https://doi.org/10.7554/elife.34532

Marshall RS, Vierstra RD (2019) Dynamic regulation of the 26S proteasome: from synthesis to degradation. Front Mol Biosci 6:40. https://doi.org/10.3389/fmolb.2019.00040

Martin A, Baker TA, Sauer RT (2008) Pore loops of the AAA + ClpX machine grip substrates to drive translocation and unfolding. Nat Struct Mol Biol 15(11):1147–1151. https://doi.org/10.1038/nsmb.1503

Martinez-Fonts K, Matouschek A (2016) A rapid and versatile method for generating proteins with defined ubiquitin chains. Biochemistry 55(12):1898–1908. https://doi.org/10.1021/acs.biochem.5b01310

Martinez-Noel G, Galligan JT, Sowa ME, Arndt V, Overton TM, Harper JW, Howley PM (2012) Identification and proteomic analysis of distinct UBE3A/E6AP protein complexes. Mol Cell Biol 32(15):3095–3106. https://doi.org/10.1128/MCB.00201-12

Martinez-Fonts K, Davis C, Tomita T, Elsasser S, Nager AR, Shi Y, Finley D, Matouschek A (2020) The proteasome 19S cap and its ubiquitin receptors provide a versatile recognition platform for substrates. Nat Commun 11(1):477. https://doi.org/10.1038/s41467-019-13906-8

Matiuhin Y, Kirkpatrick DS, Ziv I, Kim W, Dakshinamurthy A, Kleifeld O, Gygi SP, Reis N, Glickman MH (2008) Extraproteasomal Rpn10 restricts access of the polyubiquitin-binding protein Dsk2 to proteasome. Mol Cell 32(3):415–425. https://doi.org/10.1016/j.molcel.2008.10.011

Matyskiela ME, Lander GC, Martin A (2013) Conformational switching of the 26S proteasome enables substrate degradation. Nat Struct Mol Biol 20(7):781–788. https://doi.org/10.1038/nsmb.2616

Mayer RJ (2000) The meteoric rise of regulated intracellular proteolysis. Nat Rev Mol Cell Biol 1(2):145–148. https://doi.org/10.1038/35040090

Mayr J, Seemuller E, Muller SA, Engel A, Baumeister W (1998) Late events in the assembly of 20S proteasomes. J Struct Biol 124(2–3):179–188. https://doi.org/10.1006/jsbi.1998.4068

Maytal-Kivity V, Reis N, Hofmann K, Glickman MH (2002) MPN + , a putative catalytic motif found in a subset of MPN domain proteins from eukaryotes and prokaryotes, is critical for Rpn11 function. BMC Biochem 3:28. https://doi.org/10.1186/1471-2091-3-28

Mc Cormack T, Baumeister W, Grenier L, Moomaw C, Plamondon L, Pramanik B, Slaughter C, Soucy F, Stein R, Zuhl F, Dick L (1997) Active site-directed inhibitors of Rhodococcus 20 S proteasome. Kinetics and mechanism. J Biol Chem 272(42):26103–26109. https://doi.org/10.1074/jbc.272.42.26103

McCutchen-Maloney SL, Matsuda K, Shimbara N, Binns DD, Tanaka K, Slaughter CA, DeMartino GN (2000) cDNA cloning, expression, and functional characterization of PI31, a proline-rich inhibitor of the proteasome. J Biol Chem 275(24):18557–18565. https://doi.org/10.1074/jbc.M001697200

McNaught KS, Mytilineou C, Jnobaptiste R, Yabut J, Shashidharan P, Jennert P, Olanow CW (2002) Impairment of the ubiquitin-proteasome system causes dopaminergic cell death and inclusion body formation in ventral mesencephalic cultures. J Neurochem 81(2):301–306. https://doi.org/10.1046/j.1471-4159.2002.00821.x

McNaught KS, Perl DP, Brownell AL, Olanow CW (2004) Systemic exposure to proteasome inhibitors causes a progressive model of Parkinson's disease. Ann Neurol 56(1):149–162. https://doi.org/10.1002/ana.20186

Meyer H, Bug M, Bremer S (2012) Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. Nat Cell Biol 14(2):117–123. https://doi.org/10.1038/ncb2407

Meyer HJ, Rape M (2014) Enhanced protein degradation by branched ubiquitin chains. Cell 157(4):910–921. https://doi.org/10.1016/j.cell.2014.03.037

Meyer-Schwesinger C (2019) The ubiquitin-proteasome system in kidney physiology and disease. Nat Rev Nephrol 15(7):393–411. https://doi.org/10.1038/s41581-019-0148-1

Miller SB, Mogk A, Bukau B (2015) Spatially organized aggregation of misfolded proteins as cellular stress defense strategy. J Mol Biol 427(7):1564–1574. https://doi.org/10.1016/j.jmb.2015.02.006

Minis A, Rodriguez JA, Levin A, Liu K, Govek EE, Hatten ME, Steller H (2019) The proteasome regulator PI31 is required for protein homeostasis, synapse maintenance, and neuronal survival in mice. Proc Natl Acad Sci U S A 116 (49):24639–24650. https://doi.org/10.1073/pnas.1911921116

Mishto M, Bellavista E, Santoro A, Stolzing A, Ligorio C, Nacmias B, Spazzafumo L, Chiappelli M, Licastro F, Sorbi S, Pession A, Ohm T, Grune T, Franceschi C (2006) Immunoproteasome and LMP2 polymorphism in aged and Alzheimer's disease brains. Neurobiol Aging 27(1):54–66. https://doi.org/10.1016/j.neurobiolaging.2004.12.004

Monroe N, Han H, Shen PS, Sundquist WI, Hill CP (2017) Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. Elife 6. https://doi.org/10.7554/elife.24487

Moran E, Carbone F, Augusti V, Patrone F, Ballestrero A, Nencioni A (2012) Proteasome inhibitors as immunosuppressants: biological rationale and clinical experience. Semin Hematol 49(3):270–276. https://doi.org/10.1053/j.seminhematol.2012.04.004

Morris GP, Clark IA, Vissel B (2018) Questions concerning the role of amyloid-beta in the definition, aetiology and diagnosis of Alzheimer's disease. Acta Neuropathol 136(5):663–689. https://doi.org/10.1007/s00401-018-1918-8

Motosugi R, Murata S (2019) Dynamic regulation of proteasome expression. Front Mol Biosci 6:30. https://doi.org/10.3389/fmolb.2019.00030

Muchamuel T, Basler M, Aujay MA, Suzuki E, Kalim KW, Lauer C, Sylvain C, Ring ER, Shields J, Jiang J, Shwonek P, Parlati F, Demo SD, Bennett MK, Kirk CJ, Groettrup M (2009) A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. Nat Med 15(7):781–787. https://doi.org/10.1038/nm.1978

Mullane K, Williams M (2019) Preclinical models of alzheimer's disease: relevance and translational validity. Curr Protoc Pharmacol 84(1):e57. https://doi.org/10.1002/cpph.57

Muller AU, Weber-Ban E (2019) The bacterial proteasome at the core of diverse degradation pathways. Front Mol Biosci 6:23. https://doi.org/10.3389/fmolb.2019.00023

Murakami Y, Matsufuji S, Kameji T, Hayashi S, Igarashi K, Tamura T, Tanaka K, Ichihara A (1992) Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. Nature 360(6404):597–599. https://doi.org/10.1038/360597a0

Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, Tanaka K (2007) Regulation of CD8 + T cell development by thymus-specific proteasomes. Science 316(5829):1349–1353. https://doi.org/10.1126/science.1141915

Murata S, Takahama Y, Kasahara M, Tanaka K (2018) The immunoproteasome and thymoproteasome: functions, evolution and human disease. Nat Immunol 19(9):923–931. https://doi.org/10.1038/s41590-018-0186-z

Murata S, Yashiroda H, Tanaka K (2009) Molecular mechanisms of proteasome assembly. Nat Rev Mol Cell Biol 10(2):104–115. https://doi.org/10.1038/nrm2630

Nahar A, Fu X, Polovin G, Orth JD, Park S (2019) Two alternative mechanisms regulate the onset of chaperone-mediated assembly of the proteasomal ATPases. J Biol Chem 294(16):6562–6577. https://doi.org/10.1074/jbc.RA118.006298

Nakamura M, Tokunaga F, Sakata S, Iwai K (2006) Mutual regulation of conventional protein kinase C and a ubiquitin ligase complex. Biochem Biophys Res Commun 351(2):340–347. https://doi.org/10.1016/j.bbrc.2006.09.163

Nakamura Y, Nakano K, Umehara T, Kimura M, Hayashizaki Y, Tanaka A, Horikoshi M, Padmanabhan B, Yokoyama S (2007a) Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome. Structure 15(2):179–189. https://doi.org/10.1016/j.str.2006.11.015

Nakamura Y, Umehara T, Tanaka A, Horikoshi M, Padmanabhan B, Yokoyama S (2007b) Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome. Biochem Biophys Res Commun 359(3):503–509. https://doi.org/10.1016/j.bbrc.2007.05.138

Nederlof PM, Wang HR, Baumeister W (1995) Nuclear localization signals of human and Thermoplasma proteasomal alpha subunits are functional in vitro. Proc Natl Acad Sci U S A 92(26):12060–12064. https://doi.org/10.1073/pnas.92.26.12060

Nemec AA, Howell LA, Peterson AK, Murray MA, Tomko RJ Jr (2017) Autophagic clearance of proteasomes in yeast requires the conserved sorting nexin Snx4. J Biol Chem 292(52):21466–21480. https://doi.org/10.1074/jbc.M117.817999

Nemec AA, Peterson AK, Warnock JL, Reed RG, Tomko RJ, Jr. (2019) an allosteric interaction network promotes conformation state-dependent eviction of the Nas6 assembly chaperone from nascent 26S proteasomes. Cell Rep 26(2):483–495 e485. https://doi.org/10.1016/j.celrep.2018.12.042

Nencioni A, Garuti A, Schwarzenberg K, Cirmena G, Dal Bello G, Rocco I, Barbieri E, Brossart P, Patrone F, Ballestrero A (2006) Proteasome inhibitor-induced apoptosis in human monocyte-derived dendritic cells. Eur J Immunol 36(3):681–689. https://doi.org/10.1002/eji.200535298

Nickell S, Beck F, Scheres SH, Korinek A, Forster F, Lasker K, Mihalache O, Sun N, Nagy I, Sali A, Plitzko JM, Carazo JM, Mann M, Baumeister W (2009) Insights into the molecular architecture of the 26S proteasome. Proc Natl Acad Sci U S A 106(29):11943–11947. https://doi.org/10.1073/pnas.0905081106

Noda NN, Kumeta H, Nakatogawa H, Satoo K, Adachi W, Ishii J, Fujioka Y, Ohsumi Y, Inagaki F (2008) Structural basis of target recognition by Atg8/LC3 during selective autophagy. Genes Cells 13(12):1211–1218. https://doi.org/10.1111/j.1365-2443.2008.01238.x

Noda NN, Ohsumi Y, Inagaki F (2010) Atg8-family interacting motif crucial for selective autophagy. FEBS Lett 584(7):1379–1385. https://doi.org/10.1016/j.febslet.2010.01.018

Nussbaum AK, Dick TP, Keilholz W, Schirle M, Stevanovic S, Dietz K, Heinemeyer W, Groll M, Wolf DH, Huber R, Rammensee HG, Schild H (1998) Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. Proc Natl Acad Sci U S A 95(21):12504–12509. https://doi.org/10.1073/pnas.95.21.12504

Ogura T, Wilkinson AJ (2001) AAA + superfamily ATPases: common structure–diverse function. Genes Cells 6(7):575–597

Oh E, Akopian D, Rape M (2018) Principles of ubiquitin-dependent signaling. Annu Rev Cell Dev Biol 34:137–162. https://doi.org/10.1146/annurev-cellbio-100617-062802

Oh S, Hong HS, Hwang E, Sim HJ, Lee W, Shin SJ, Mook-Jung I (2005) Amyloid peptide attenuates the proteasome activity in neuronal cells. Mech Ageing Dev 126(12):1292–1299. https://doi.org/10.1016/j.mad.2005.07.006

Ohno A, Jee J, Fujiwara K, Tenno T, Goda N, Tochio H, Kobayashi H, Hiroaki H, Shirakawa M (2005) Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition. Structure 13(4):521–532. https://doi.org/10.1016/j.str.2005.01.011

Ohsumi Y (2001) Molecular dissection of autophagy: two ubiquitin-like systems. Nat Rev Mol Cell Biol 2(3):211–216. https://doi.org/10.1038/35056522

Oinonen C, Rouvinen J (2000) Structural comparison of Ntn-hydrolases. Protein Sci 9(12):2329–2337. https://doi.org/10.1110/ps.9.12.2329

Olivares AO, Baker TA, Sauer RT (2018) Mechanical protein unfolding and degradation. Annu Rev Physiol 80:413–429. https://doi.org/10.1146/annurev-physiol-021317-121303

Olszewski MM, Williams C, Dong KC, Martin A (2019) The Cdc48 unfoldase prepares well-folded protein substrates for degradation by the 26S proteasome. Commun Biol 2:29. https://doi.org/10.1038/s42003-019-0283-z

Orlowski M (1990) The multicatalytic proteinase complex, a major extralysosomal proteolytic system. Biochemistry 29(45):10289–10297. https://doi.org/10.1021/bi00497a001

Orlowski RZ, Eswara JR, Lafond-Walker A, Grever MR, Orlowski M, Dang CV (1998) Tumor growth inhibition induced in a murine model of human Burkitt's lymphoma by a proteasome inhibitor. Cancer Res 58(19):4342–4348

Orlowski RZ, Stinchcombe TE, Mitchell BS, Shea TC, Baldwin AS, Stahl S, Adams J, Esseltine DL, Elliott PJ, Pien CS, Guerciolini R, Anderson JK, Depcik-Smith ND, Bhagat R, Lehman MJ, Novick SC, O'Connor OA, Soignet SL (2002) Phase I trial of the proteasome inhibitor PS-341 in patients with refractory hematologic malignancies. J Clin Oncol 20(22):4420–4427. https://doi.org/10.1200/JCO.2002.01.133

Ortega Z, Diaz-Hernandez M, Lucas JJ (2007) Is the ubiquitin-proteasome system impaired in Huntington's disease? Cell Mol Life Sci 64(17):2245–2257. https://doi.org/10.1007/s00018-007-7222-8

Ortega Z, Lucas JJ (2014) Ubiquitin-proteasome system involvement in Huntington's disease. Front Mol Neurosci 7:77. https://doi.org/10.3389/fnmol.2014.00077

Ortolan TG, Tongaonkar P, Lambertson D, Chen L, Schauber C, Madura K (2000) The DNA repair protein rad23 is a negative regulator of multi-ubiquitin chain assembly. Nat Cell Biol 2(9):601–608. https://doi.org/10.1038/35023547

Osmulski PA, Hochstrasser M, Gaczynska M (2009) A tetrahedral transition state at the active sites of the 20S proteasome is coupled to opening of the alpha-ring channel. Structure 17(8):1137–1147. https://doi.org/10.1016/j.str.2009.06.011

Osorio FG, Soria-Valles C, Santiago-Fernandez O, Bernal T, Mittelbrunn M, Colado E, Rodriguez F, Bonzon-Kulichenko E, Vazquez J, Porta-de-la-Riva M, Ceron J, Fueyo A, Li J, Green AR, Freije JM, Lopez-Otin C (2016) Loss of the proteostasis factor AIRAPL causes myeloid transformation by deregulating IGF-1 signaling. Nat Med 22(1):91–96. https://doi.org/10.1038/nm.4013

Owsianik G, Balzi l L, Ghislain M (2002) Control of 26S proteasome expression by transcription factors regulating multidrug resistance in Saccharomyces cerevisiae. Mol Microbiol 43(5):1295–1308. https://doi.org/10.1046/j.1365-2958.2002.02823.x

Pack CG, Yukii H, Toh-e A, Kudo T, Tsuchiya H, Kaiho A, Sakata E, Murata S, Yokosawa H, Sako Y, Baumeister W, Tanaka K, Saeki Y (2014) Quantitative live-cell imaging reveals spatio-temporal dynamics and cytoplasmic assembly of the 26S proteasome. Nat Commun 5:3396. https://doi.org/10.1038/ncomms4396

Padmanabhan A, Vuong SA, Hochstrasser M (2016) Assembly of an evolutionarily conserved alternative proteasome isoform in human cells. Cell Rep 14(12):2962–2974. https://doi.org/10.1016/j.celrep.2016.02.068

Pagan J, Seto T, Pagano M, Cittadini A (2013) Role of the ubiquitin proteasome system in the heart. Circ Res 112(7):1046–1058. https://doi.org/10.1161/CIRCRESAHA.112.300521

Pagano M, Tam SW, Theodoras AM, Beer-Romero P, Del Sal G, Chau V, Yew PR, Draetta GF, Rolfe M (1995) Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. Science 269(5224):682–685. https://doi.org/10.1126/science.7624798

Pamnani V, Tamura T, Lupas A, Peters J, Cejka Z, Ashraf W, Baumeister W (1997) Cloning, sequencing and expression of VAT, a CDC48/p97 ATPase homologue from the archaeon Thermoplasma acidophilum. FEBS Lett 404(2–3):263–268. https://doi.org/10.1016/s0014-5793(97)00138-5

Pante N, Kann M (2002) Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. Mol Biol Cell 13(2):425–434. https://doi.org/10.1091/mbc.01-06-0308

Papadopoulos C, Kirchner P, Bug M, Grum D, Koerver L, Schulze N, Poehler R, Dressler A, Fengler S, Arhzaouy K, Lux V, Ehrmann M, Weihl CC, Meyer H (2017) VCP/p97 cooperates

with YOD1, UBXD1 and PLAA to drive clearance of ruptured lysosomes by autophagy. EMBO J 36(2):135–150. https://doi.org/10.15252/embj.201695148
Park S, Kim W, Tian G, Gygi SP, Finley D (2011) Structural defects in the regulatory particle-core particle interface of the proteasome induce a novel proteasome stress response. J Biol Chem 286(42):36652–36666. https://doi.org/10.1074/jbc.M111.285924
Park S, Li X, Kim HM, Singh CR, Tian G, Hoyt MA, Lovell S, Battaile KP, Zolkiewski M, Coffino P, Roelofs J, Cheng Y, Finley D (2013) Reconfiguration of the proteasome during chaperone-mediated assembly. Nature 497(7450):512–516. https://doi.org/10.1038/nature12123
Park S, Roelofs J, Kim W, Robert J, Schmidt M, Gygi SP, Finley D (2009) Hexameric assembly of the proteasomal ATPases is templated through their C termini. Nature 459(7248):866–870. https://doi.org/10.1038/nature08065
Park S, Tian G, Roelofs J, Finley D (2010) Assembly manual for the proteasome regulatory particle: the first draft. Biochem Soc Trans 38(Pt 1):6–13. https://doi.org/10.1042/BST0380006
Parzych KR, Klionsky DJ (2019) Vacuolar hydrolysis and efflux: current knowledge and unanswered questions. Autophagy 15(2):212–227. https://doi.org/10.1080/15548627.2018.1545821
Patel SK, Indig FE, Olivieri N, Levine ND, Latterich M (1998) Organelle membrane fusion: a novel function for the syntaxin homolog Ufe1p in ER membrane fusion. Cell 92(5):611–620. https://doi.org/10.1016/s0092-8674(00)81129-0
Pathare GR, Nagy I, Bohn S, Unverdorben P, Hubert A, Korner R, Nickell S, Lasker K, Sali A, Tamura T, Nishioka T, Forster F, Baumeister W, Bracher A (2012) The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. Proc Natl Acad Sci U S A 109(1):149–154. https://doi.org/10.1073/pnas.1117648108
Pathare GR, Nagy I, Sledz P, Anderson DJ, Zhou HJ, Pardon E, Steyaert J, Forster F, Bracher A, Baumeister W (2014) Crystal structure of the proteasomal deubiquitylation module Rpn8-Rpn11. Proc Natl Acad Sci U S A 111(8):2984–2989. https://doi.org/10.1073/pnas.1400546111
Perez VI, Buffenstein R, Masamsetti V, Leonard S, Salmon AB, Mele J, Andziak B, Yang T, Edrey Y, Friguet B, Ward W, Richardson A, Chaudhuri A (2009) Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. Proc Natl Acad Sci U S A 106(9):3059–3064. https://doi.org/10.1073/pnas.0809620106
Peters JM (2006) The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol 7(9):644–656. https://doi.org/10.1038/nrm1988
Peters JM, Cejka Z, Harris JR, Kleinschmidt JA, Baumeister W (1993) Structural features of the 26 S proteasome complex. J Mol Biol 234(4):932–937. https://doi.org/10.1006/jmbi.1993.1646
Peters LZ, Karmon O, David-Kadoch G, Hazan R, Yu T, Glickman MH, Ben-Aroya S (2015) The protein quality control machinery regulates its misassembled proteasome subunits. PLoS Genet 11(4):e1005178. https://doi.org/10.1371/journal.pgen.1005178
Peters LZ, Karmon O, Miodownik S, Ben-Aroya S (2016) Proteasome storage granules are transiently associated with the insoluble protein deposit in Saccharomyces cerevisiae. J Cell Sci 129(6):1190–1197. https://doi.org/10.1242/jcs.179648
Peth A, Besche HC, Goldberg AL (2009) Ubiquitinated proteins activate the proteasome by binding to Usp14/Ubp6, which causes 20S gate opening. Mol Cell 36(5):794–804. https://doi.org/10.1016/j.molcel.2009.11.015
Peth A, Kukushkin N, Bosse M, Goldberg AL (2013a) Ubiquitinated proteins activate the proteasomal ATPases by binding to Usp14 or Uch37 homologs. J Biol Chem 288(11):7781–7790. https://doi.org/10.1074/jbc.M112.441907
Peth A, Nathan JA, Goldberg AL (2013b) The ATP costs and time required to degrade ubiquitinated proteins by the 26 S proteasome. J Biol Chem 288(40):29215–29222. https://doi.org/10.1074/jbc.M113.482570
Peth A, Uchiki T, Goldberg AL (2010) ATP-dependent steps in the binding of ubiquitin conjugates to the 26S proteasome that commit to degradation. Mol Cell 40(4):671–681. https://doi.org/10.1016/j.molcel.2010.11.002
Pickart CM (2000) Ubiquitin in chains. Trends Biochem Sci 25(11):544–548. https://doi.org/10.1016/s0968-0004(00)01681-9

Piwko W, Jentsch S (2006) Proteasome-mediated protein processing by bidirectional degradation initiated from an internal site. Nat Struct Mol Biol 13(8):691–697. https://doi.org/10.1038/nsmb1122

Powell SR, Herrmann J, Lerman A, Patterson C, Wang X (2012) The ubiquitin-proteasome system and cardiovascular disease. Prog Mol Biol Transl Sci 109:295–346. https://doi.org/10.1016/B978-0-12-397863-9.00009-2

Powell SR, Samuel SM, Wang P, Divald A, Thirunavukkarasu M, Koneru S, Wang X, Maulik N (2008) Upregulation of myocardial 11S-activated proteasome in experimental hyperglycemia. J Mol Cell Cardiol 44(3):618–621. https://doi.org/10.1016/j.yjmcc.2007.12.009

Prakash S, Inobe T, Hatch AJ, Matouschek A (2009) Substrate selection by the proteasome during degradation of protein complexes. Nat Chem Biol 5(1):29–36. https://doi.org/10.1038/nchembio.130

Prakash S, Tian L, Ratliff KS, Lehotzky RE, Matouschek A (2004) An unstructured initiation site is required for efficient proteasome-mediated degradation. Nat Struct Mol Biol 11(9):830–837. https://doi.org/10.1038/nsmb814

Puchades C, Ding B, Song A, Wiseman RL, Lander GC, Glynn SE (2019) Unique structural features of the mitochondrial AAA + Protease AFG3L2 reveal the molecular basis for activity in health and disease. Mol Cell. https://doi.org/10.1016/j.molcel.2019.06.016

Puchades C, Rampello AJ, Shin M, Giuliano CJ, Wiseman RL, Glynn SE, Lander GC (2017) Structure of the mitochondrial inner membrane AAA + protease YME1 gives insight into substrate processing. Science 358(6363). https://doi.org/10.1126/science.aao0464

Puchades C, Sandate CR, Lander GC (2020) The molecular principles governing the activity and functional diversity of AAA + proteins. Nat Rev Mol Cell Biol 21(1):43–58. https://doi.org/10.1038/s41580-019-0183-6

Qin Y, Zhou MT, Hu MM, Hu YH, Zhang J, Guo L, Zhong B, Shu HB (2014) RNF26 temporally regulates virus-triggered type I interferon induction by two distinct mechanisms. PLoS Pathog 10(9):e1004358. https://doi.org/10.1371/journal.ppat.1004358

Qiu XB, Ouyang SY, Li CJ, Miao S, Wang L, Goldberg AL (2006) hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. EMBO J 25(24):5742–5753. https://doi.org/10.1038/sj.emboj.7601450

Raasi S, Pickart CM (2003) Rad23 ubiquitin-associated domains (UBA) inhibit 26 S proteasome-catalyzed proteolysis by sequestering lysine 48-linked polyubiquitin chains. J Biol Chem 278(11):8951–8959. https://doi.org/10.1074/jbc.m212841200

Raasi S, Varadan R, Fushman D, Pickart CM (2005) Diverse polyubiquitin interaction properties of ubiquitin-associated domains. Nat Struct Mol Biol 12(8):708–714. https://doi.org/10.1038/nsmb962

Rabinovich E, Kerem A, Frohlich KU, Diamant N, Bar-Nun S (2002) AAA-ATPase p97/Cdc48p, a cytosolic chaperone required for endoplasmic reticulum-associated protein degradation. Mol Cell Biol 22(2):626–634. https://doi.org/10.1128/mcb.22.2.626-634.2002

Rabl J, Smith DM, Yu Y, Chang SC, Goldberg AL, Cheng Y (2008) Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. Mol Cell 30(3):360–368. https://doi.org/10.1016/j.molcel.2008.03.004

Radhakrishnan SK, Lee CS, Young P, Beskow A, Chan JY, Deshaies RJ (2010) Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. Mol Cell 38(1):17–28. https://doi.org/10.1016/j.molcel.2010.02.029

Rahighi S, Braunstein I, Ternette N, Kessler B, Kawasaki M, Kato R, Matsui T, Weiss TM, Stanhill A, Wakatsuki S (2016) Selective binding of AIRAPL Tandem UIMs to Lys48-linked tri-ubiquitin chains. Structure 24(3):412–422. https://doi.org/10.1016/j.str.2015.12.017

Ramachandran KV, Fu JM, Schaffer TB, Na CH, Delannoy M, Margolis SS (2018) Activity-dependent degradation of the nascentome by the neuronal membrane proteasome. Mol Cell 71(1):169–177 e166. https://doi.org/10.1016/j.molcel.2018.06.013

Ramachandran KV, Margolis SS (2017) A mammalian nervous-system-specific plasma membrane proteasome complex that modulates neuronal function. Nat Struct Mol Biol 24(4):419–430. https://doi.org/10.1038/nsmb.3389

Ramadan K, Bruderer R, Spiga FM, Popp O, Baur T, Gotta M, Meyer HH (2007) Cdc48/p97 promotes reformation of the nucleus by extracting the kinase Aurora B from chromatin. Nature 450(7173):1258–1262. https://doi.org/10.1038/nature06388

Ramanathan HN, Ye Y (2012) Cellular strategies for making monoubiquitin signals. Crit Rev Biochem Mol Biol 47(1):17–28. https://doi.org/10.3109/10409238.2011.620943

Ramos PC, Hockendorff J, Johnson ES, Varshavsky A, Dohmen RJ (1998) Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. Cell 92(4):489–499. https://doi.org/10.1016/s0092-8674(00)80942-3

Realini C, Dubiel W, Pratt G, Ferrell K, Rechsteiner M (1994) Molecular cloning and expression of a gamma-interferon-inducible activator of the multicatalytic protease. J Biol Chem 269(32):20727–20732

Rechsteiner M, Hill CP (2005) Mobilizing the proteolytic machine: cell biological roles of proteasome activators and inhibitors. Trends Cell Biol 15(1):27–33. https://doi.org/10.1016/j.tcb.2004.11.003

Reggiori F, Klionsky DJ (2013) Autophagic processes in yeast: mechanism, machinery and regulation. Genetics 194(2):341–361. https://doi.org/10.1534/genetics.112.149013

Reits EA, Benham AM, Plougastel B, Neefjes J, Trowsdale J (1997) Dynamics of proteasome distribution in living cells. EMBO J 16(20):6087–6094. https://doi.org/10.1093/emboj/16.20.6087

Religa TL, Sprangers R, Kay LE (2010) Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science 328(5974):98–102. https://doi.org/10.1126/science.1184991

Remondelli P, Renna M (2017) The endoplasmic reticulum unfolded protein response in neurodegenerative disorders and its potential therapeutic significance. Front Mol Neurosci 10:187. https://doi.org/10.3389/fnmol.2017.00187

Renaud JP, Chari A, Ciferri C, Liu WT, Remigy HW, Stark H, Wiesmann C (2018) Cryo-EM in drug discovery: achievements, limitations and prospects. Nat Rev Drug Discov 17(7):471–492. https://doi.org/10.1038/nrd.2018.77

Richly H, Rape M, Braun S, Rumpf S, Hoege C, Jentsch S (2005) A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. Cell 120(1):73–84. https://doi.org/10.1016/j.cell.2004.11.013

Ricoult SJ, Manning BD (2013) The multifaceted role of mTORC1 in the control of lipid metabolism. EMBO Rep 14(3):242–251. https://doi.org/10.1038/embor.2013.5

Riedinger C, Boehringer J, Trempe JF, Lowe ED, Brown NR, Gehring K, Noble ME, Gordon C, Endicott JA (2010) Structure of Rpn10 and its interactions with polyubiquitin chains and the proteasome subunit Rpn12. J Biol Chem 285(44):33992–34003. https://doi.org/10.1074/jbc.M110.134510

Ripstein ZA, Huang R, Augustyniak R, Kay LE, Rubinstein JL (2017) Structure of a AAA + unfoldase in the process of unfolding substrate. Elife 6. https://doi.org/10.7554/elife.25754

Ripstein ZA, Vahidi S, Houry WA, Rubinstein JL, Kay LE (2020) A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. Elife 9. https://doi.org/10.7554/elife.52158

Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, Hwang D, Goldberg AL (1994) Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell 78(5):761–771. https://doi.org/10.1016/s0092-8674(94)90462-6

Roelofs J, Park S, Haas W, Tian G, McAllister FE, Huo Y, Lee BH, Zhang F, Shi Y, Gygi SP, Finley D (2009) Chaperone-mediated pathway of proteasome regulatory particle assembly. Nature 459(7248):861–865. https://doi.org/10.1038/nature08063

Roeten MSF, Cloos J, Jansen G (2018) Positioning of proteasome inhibitors in therapy of solid malignancies. Cancer Chemother Pharmacol 81(2):227–243. https://doi.org/10.1007/s00280-017-3489-0

Rogov V, Dotsch V, Johansen T, Kirkin V (2014) Interactions between autophagy receptors and ubiquitin-like proteins form the molecular basis for selective autophagy. Mol Cell 53(2):167–178. https://doi.org/10.1016/j.molcel.2013.12.014

Rogov VV, Stolz A, Ravichandran AC, Rios-Szwed DO, Suzuki H, Kniss A, Lohr F, Wakatsuki S, Dotsch V, Dikic I, Dobson RC, McEwan DG (2017) Structural and functional analysis of the GABARAP interaction motif (GIM). EMBO Rep 18(8):1382–1396. https://doi.org/10.15252/embr.201643587

Romano R, Palamaro L, Fusco A, Giardino G, Gallo V, Del Vecchio L, Pignata C (2013) FOXN1: a master regulator gene of thymic epithelial development program. Front Immunol 4:187. https://doi.org/10.3389/fimmu.2013.00187

Rosenzweig R, Osmulski PA, Gaczynska M, Glickman MH (2008) The central unit within the 19S regulatory particle of the proteasome. Nat Struct Mol Biol 15(6):573–580. https://doi.org/10.1038/nsmb.1427

Rousseau A, Bertolotti A (2016) An evolutionarily conserved pathway controls proteasome homeostasis. Nature 536(7615):184–189. https://doi.org/10.1038/nature18943

Rousseau A, Bertolotti A (2018) Regulation of proteasome assembly and activity in health and disease. Nat Rev Mol Cell Biol 19(11):697–712. https://doi.org/10.1038/s41580-018-0040-z

Rubinsztein DC (2006) The roles of intracellular protein-degradation pathways in neurodegeneration. Nature 443(7113):780–786. https://doi.org/10.1038/nature05291

Ruggeri B, Miknyoczki S, Dorsey B, Hui AM (2009) The development and pharmacology of proteasome inhibitors for the management and treatment of cancer. Adv Pharmacol 57:91–135. https://doi.org/10.1016/S1054-3589(08)57003-7

Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE (2010) The proteasome antechamber maintains substrates in an unfolded state. Nature 467(7317):868–871. https://doi.org/10.1038/nature09444

Russell SJ, Steger KA, Johnston SA (1999) Subcellular localization, stoichiometry, and protein levels of 26 S proteasome subunits in yeast. J Biol Chem 274(31):21943–21952. https://doi.org/10.1074/jbc.274.31.21943

Sadre-Bazzaz K, Whitby FG, Robinson H, Formosa T, Hill CP (2010) Structure of a Blm10 complex reveals common mechanisms for proteasome binding and gate opening. Mol Cell 37(5):728–735. https://doi.org/10.1016/j.molcel.2010.02.002

Saeki Y (2017) Ubiquitin recognition by the proteasome. J Biochem 161(2):113–124. https://doi.org/10.1093/jb/mvw091

Saeki Y, Kudo T, Sone T, Kikuchi Y, Yokosawa H, Toh-e A, Tanaka K (2009a) Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. EMBO J 28(4):359–371. https://doi.org/10.1038/emboj.2008.305

Saeki Y, Toh EA, Kudo T, Kawamura H, Tanaka K (2009b) Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle. Cell 137(5):900–913. https://doi.org/10.1016/j.cell.2009.05.005

Saez I, Vilchez D (2014) The mechanistic links between proteasome activity, aging and age-related diseases. Curr Genomics 15(1):38–51. https://doi.org/10.2174/1389202915011403061133344

Sahtoe DD, van Dijk WJ, El Oualid F, Ekkebus R, Ovaa H, Sixma TK (2015) Mechanism of UCH-L5 activation and inhibition by DEUBAD domains in RPN13 and INO80G. Mol Cell 57(5):887–900. https://doi.org/10.1016/j.molcel.2014.12.039

Sakata E, Bohn S, Mihalache O, Kiss P, Beck F, Nagy I, Nickell S, Tanaka K, Saeki Y, Forster F, Baumeister W (2012) Localization of the proteasomal ubiquitin receptors Rpn10 and Rpn13 by electron cryomicroscopy. Proc Natl Acad Sci U S A 109(5):1479–1484. https://doi.org/10.1073/pnas.1119394109

Sakata E, Yamaguchi Y, Kurimoto E, Kikuchi J, Yokoyama S, Yamada S, Kawahara H, Yokosawa H, Hattori N, Mizuno Y, Tanaka K, Kato K (2003) Parkin binds the Rpn10 subunit of 26S proteasomes

through its ubiquitin-like domain. EMBO Rep 4(3):301–306. https://doi.org/10.1038/sj.embor.embor764

Salmon AB, Leonard S, Masamsetti V, Pierce A, Podlutsky AJ, Podlutskaya N, Richardson A, Austad SN, Chaudhuri AR (2009) The long lifespan of two bat species is correlated with resistance to protein oxidation and enhanced protein homeostasis. FASEB J 23(7):2317–2326. https://doi.org/10.1096/fj.08-122523

Samant RS, Livingston CM, Sontag EM, Frydman J (2018) Distinct proteostasis circuits cooperate in nuclear and cytoplasmic protein quality control. Nature 563(7731):407–411. https://doi.org/10.1038/s41586-018-0678-x

Sandstrom A, Mitchell PS, Goers L, Mu EW, Lesser CF, Vance RE (2019) Functional degradation: a mechanism of NLRP1 inflammasome activation by diverse pathogen enzymes. Science 364(6435). https://doi.org/10.1126/science.aau1330

Sato Y, Goto E, Shibata Y, Kubota Y, Yamagata A, Goto-Ito S, Kubota K, Inoue J, Takekawa M, Tokunaga F, Fukai S (2015) Structures of CYLD USP with Met1- or Lys63-linked diubiquitin reveal mechanisms for dual specificity. Nat Struct Mol Biol 22(3):222–229. https://doi.org/10.1038/nsmb.2970

Sato Y, Yoshikawa A, Yamagata A, Mimura H, Yamashita M, Ookata K, Nureki O, Iwai K, Komada M, Fukai S (2008) Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains. Nature 455(7211):358–362. https://doi.org/10.1038/nature07254

Satoh K, Sasajima H, Nyoumura KI, Yokosawa H, Sawada H (2001) Assembly of the 26S proteasome is regulated by phosphorylation of the p45/Rpt6 ATPase subunit. Biochemistry 40(2):314–319. https://doi.org/10.1021/bi001815n

Sauer RT, Baker TA (2011) AAA + proteases: ATP-fueled machines of protein destruction. Annu Rev Biochem 80:587–612. https://doi.org/10.1146/annurev-biochem-060408-172623

Savulescu AF, Glickman MH (2011) Proteasome activator 200: the heat is on. Mol Cell Proteomics 10(5):R110 006890. https://doi.org/10.1074/mcp.r110.006890

Savulescu AF, Shorer H, Kleifeld O, Cohen I, Gruber R, Glickman MH, Harel A (2011) Nuclear import of an intact preassembled proteasome particle. Mol Biol Cell 22(6):880–891. https://doi.org/10.1091/mbc.E10-07-0595

Scapin G, Potter CS, Carragher B (2018) Cryo-EM for small molecules discovery, design, understanding, and application. Cell Chem Biol 25(11):1318–1325. https://doi.org/10.1016/j.chembiol.2018.07.006

Schauber C, Chen L, Tongaonkar P, Vega I, Lambertson D, Potts W, Madura K (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway. Nature 391(6668):715–718. https://doi.org/10.1038/35661

Scherer DC, Brockman JA, Chen Z, Maniatis T, Ballard DW (1995) Signal-induced degradation of I kappa B alpha requires site-specific ubiquitination. Proc Natl Acad Sci U S A 92(24):11259–11263. https://doi.org/10.1073/pnas.92.24.11259

Schlossarek S, Carrier L (2011) The ubiquitin-proteasome system in cardiomyopathies. Curr Opin Cardiol 26(3):190–195. https://doi.org/10.1097/HCO.0b013e32834598fe

Schmidt M, Haas W, Crosas B, Santamaria PG, Gygi SP, Walz T, Finley D (2005) The HEAT repeat protein Blm10 regulates the yeast proteasome by capping the core particle. Nat Struct Mol Biol 12(4):294–303. https://doi.org/10.1038/nsmb914

Schmidtke G, Kraft R, Kostka S, Henklein P, Frommel C, Lowe J, Huber R, Kloetzel PM, Schmidt M (1996) Analysis of mammalian 20S proteasome biogenesis: the maturation of beta-subunits is an ordered two-step mechanism involving autocatalysis. EMBO J 15(24):6887–6898

Schrader J, Henneberg F, Mata RA, Tittmann K, Schneider TR, Stark H, Bourenkov G, Chari A (2016) The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. Science 353(6299):594–598. https://doi.org/10.1126/science.aaf8993

Schreiner P, Chen X, Husnjak K, Randles L, Zhang N, Elsasser S, Finley D, Dikic I, Walters KJ, Groll M (2008) Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. Nature 453(7194):548–552. https://doi.org/10.1038/nature06924

Schweitzer A, Aufderheide A, Rudack T, Beck F, Pfeifer G, Plitzko JM, Sakata E, Schulten K, Forster F, Baumeister W (2016) Structure of the human 26S proteasome at a resolution of 3.9 A. Proc Natl Acad Sci U S A 113(28):7816–7821. https://doi.org/10.1073/pnas.1608050113

Schwertman P, Bekker-Jensen S, Mailand N (2016) Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. Nat Rev Mol Cell Biol 17(6):379–394. https://doi.org/10.1038/nrm.2016.58

Seemuller E, Lupas A, Baumeister W (1996) Autocatalytic processing of the 20S proteasome. Nature 382(6590):468–471. https://doi.org/10.1038/382468a0

Seemuller E, Lupas A, Stock D, Lowe J, Huber R, Baumeister W (1995a) Proteasome from Thermoplasma acidophilum: a threonine protease. Science 268(5210):579–582. https://doi.org/10.1126/science.7725107

Seemuller E, Lupas A, Zuhl F, Zwickl P, Baumeister W (1995b) The proteasome from Thermoplasma acidophilum is neither a cysteine nor a serine protease. FEBS Lett 359(2–3):173–178. https://doi.org/10.1016/0014-5793(95)00036-9

Selkoe DJ (2003) Folding proteins in fatal ways. Nature 426(6968):900–904. https://doi.org/10.1038/nature02264

Shabek N, Herman-Bachinsky Y, Buchsbaum S, Lewinson O, Haj-Yahya M, Hejjaoui M, Lashuel HA, Sommer T, Brik A, Ciechanover A (2012) The size of the proteasomal substrate determines whether its degradation will be mediated by mono- or polyubiquitylation. Mol Cell 48(1):87–97. https://doi.org/10.1016/j.molcel.2012.07.011

Shah SA, Potter MW, McDade TP, Ricciardi R, Perugini RA, Elliott PJ, Adams J, Callery MP (2001) 26S proteasome inhibition induces apoptosis and limits growth of human pancreatic cancer. J Cell Biochem 82(1):110–122. https://doi.org/10.1002/jcb.1150

Sharif H, Wang L, Wang WL, Magupalli VG, Andreeva L, Qiao Q, Hauenstein AV, Wu Z, Nunez G, Mao Y, Wu H (2019) Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome. Nature 570(7761):338–343. https://doi.org/10.1038/s41586-019-1295-z

Sharipo A, Imreh M, Leonchiks A, Imreh S, Masucci MG (1998) A minimal glycine-alanine repeat prevents the interaction of ubiquitinated I kappaB alpha with the proteasome: a new mechanism for selective inhibition of proteolysis. Nat Med 4(8):939–944. https://doi.org/10.1038/nm0898-939

Shi Y, Chen X, Elsasser S, Stocks BB, Tian G, Lee BH, Shi Y, Zhang N, de Poot SA, Tuebing F, Sun S, Vannoy J, Tarasov SG, Engen JR, Finley D, Walters KJ (2016) Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. Science 351(6275). https://doi.org/10.1126/science.aad9421

Shibahara T, Kawasaki H, Hirano H (2002) Identification of the 19S regulatory particle subunits from the rice 26S proteasome. Eur J Biochem 269(5):1474–1483. https://doi.org/10.1046/j.1432-1033.2002.02792.x

Shimizu Y, Taraborrelli L, Walczak H (2015) Linear ubiquitination in immunity. Immunol Rev 266(1):190–207. https://doi.org/10.1111/imr.12309

Shirozu R, Yashiroda H, Murata S (2015) Identification of minimum Rpn4-responsive elements in genes related to proteasome functions. FEBS Lett 589(8):933–940. https://doi.org/10.1016/j.febslet.2015.02.025

Shrestha RK, Ronau JA, Davies CW, Guenette RG, Strieter ER, Paul LN, Das C (2014) Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. Biochemistry 53(19):3199–3217. https://doi.org/10.1021/bi5003162

Shukla SK, Rafiq K (2019) Proteasome biology and therapeutics in cardiac diseases. Transl Res 205:64–76. https://doi.org/10.1016/j.trsl.2018.09.003

Sijts A, Sun Y, Janek K, Kral S, Paschen A, Schadendorf D, Kloetzel PM (2002) The role of the proteasome activator PA28 in MHC class I antigen processing. Mol Immunol 39(3–4):165–169. https://doi.org/10.1016/s0161-5890(02)00099-8

Sijts EJ, Kloetzel PM (2011) The role of the proteasome in the generation of MHC class I ligands and immune responses. Cell Mol Life Sci 68(9):1491–1502. https://doi.org/10.1007/s00018-011-0657-y

Silva GM, Finley D, Vogel C (2015) K63 polyubiquitination is a new modulator of the oxidative stress response. Nat Struct Mol Biol 22(2):116–123. https://doi.org/10.1038/nsmb.2955

Silva GM, Netto LE, Simoes V, Santos LF, Gozzo FC, Demasi MA, Oliveira CL, Bicev RN, Klitzke CF, Sogayar MC, Demasi M (2012) Redox control of 20S proteasome gating. Antioxid Redox Signal 16(11):1183–1194. https://doi.org/10.1089/ars.2011.4210

Skrott Z, Mistrik M, Andersen KK, Friis S, Majera D, Gursky J, Ozdian T, Bartkova J, Turi Z, Moudry P, Kraus M, Michalova M, Vaclavkova J, Dzubak P, Vrobel I, Pouckova P, Sedlacek J, Miklovicova A, Kutt A, Li J, Mattova J, Driessen C, Dou QP, Olsen J, Hajduch M, Cvek B, Deshaies RJ, Bartek J (2017) Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4. Nature 552(7684):194–199. https://doi.org/10.1038/nature25016

Sledz P, Unverdorben P, Beck F, Pfeifer G, Schweitzer A, Forster F, Baumeister W (2013) Structure of the 26S proteasome with ATP-gammaS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. Proc Natl Acad Sci U S A 110(18):7264–7269. https://doi.org/10.1073/pnas.1305782110

Sloper-Mould KE, Jemc JC, Pickart CM, Hicke L (2001) Distinct functional surface regions on ubiquitin. J Biol Chem 276(32):30483–30489. https://doi.org/10.1074/jbc.M103248200

Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MS (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph 14(6):354–360, 376

Smith DM, Benaroudj N, Goldberg A (2006) Proteasomes and their associated ATPases: a destructive combination. J Struct Biol 156(1):72–83. https://doi.org/10.1016/j.jsb.2006.04.012

Smith DM, Chang SC, Park S, Finley D, Cheng Y, Goldberg AL (2007) Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. Mol Cell 27(5):731–744. https://doi.org/10.1016/j.molcel.2007.06.033

Smith DM, Fraga H, Reis C, Kafri G, Goldberg AL (2011) ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. Cell 144(4):526–538. https://doi.org/10.1016/j.cell.2011.02.005

Smith DM, Kafri G, Cheng Y, Ng D, Walz T, Goldberg AL (2005) ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins. Mol Cell 20(5):687–698. https://doi.org/10.1016/j.molcel.2005.10.019

Smith HL, Mallucci GR (2016) The unfolded protein response: mechanisms and therapy of neurodegeneration. Brain 139(Pt 8):2113–2121. https://doi.org/10.1093/brain/aww101

Snoberger A, Brettrager EJ, Smith DM (2018) Conformational switching in the coiled-coil domains of a proteasomal ATPase regulates substrate processing. Nat Commun 9(1):2374. https://doi.org/10.1038/s41467-018-04731-6

Song EJ, Werner SL, Neubauer J, Stegmeier F, Aspden J, Rio D, Harper JW, Elledge SJ, Kirschner MW, Rape M (2010) The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes Dev 24(13):1434–1447. https://doi.org/10.1101/gad.1925010

Song Y, Park PMC, Wu L, Ray A, Picaud S, Li D, Wimalasena VK, Du T, Filippakopoulos P, Anderson KC, Qi J, Chauhan D (2019) Development and preclinical validation of a novel covalent ubiquitin receptor Rpn13 degrader in multiple myeloma. Leukemia 33(11):2685–2694. https://doi.org/10.1038/s41375-019-0467-z

Song Y, Ray A, Li S, Das DS, Tai YT, Carrasco RD, Chauhan D, Anderson KC (2016) Targeting proteasome ubiquitin receptor Rpn13 in multiple myeloma. Leukemia 30(9):1877–1886. https://doi.org/10.1038/leu.2016.97

Specht S, Miller SB, Mogk A, Bukau B (2011) Hsp42 is required for sequestration of protein aggregates into deposition sites in Saccharomyces cerevisiae. J Cell Biol 195(4):617–629. https://doi.org/10.1083/jcb.201106037

Spence J, Gali RR, Dittmar G, Sherman F, Karin M, Finley D (2000) Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. Cell 102(1):67–76. https://doi.org/10.1016/s0092-8674(00)00011-8

Spencer E, Jiang J, Chen ZJ (1999) Signal-induced ubiquitination of IkappaBalpha by the F-box protein Slimb/beta-TrCP. Genes Dev 13(3):284–294. https://doi.org/10.1101/gad.13.3.284

Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445(7128):618–622. https://doi.org/10.1038/nature05512

Sprangers R, Li X, Mao X, Rubinstein JL, Schimmer AD, Kay LE (2008) TROSY-based NMR evidence for a novel class of 20S proteasome inhibitors. Biochemistry 47(26):6727–6734. https://doi.org/10.1021/bi8005913

Stadtmueller BM, Ferrell K, Whitby FG, Heroux A, Robinson H, Myszka DG, Hill CP (2010) Structural models for interactions between the 20S proteasome and its PAN/19S activators. J Biol Chem 285(1):13–17. https://doi.org/10.1074/jbc.C109.070425

Stadtmueller BM, Hill CP (2011) Proteasome activators. Mol Cell 41(1):8–19. https://doi.org/10.1016/j.molcel.2010.12.020

Stadtmueller BM, Kish-Trier E, Ferrell K, Petersen CN, Robinson H, Myszka DG, Eckert DM, Formosa T, Hill CP (2012) Structure of a proteasome Pba1-Pba2 complex: implications for proteasome assembly, activation, and biological function. J Biol Chem 287(44):37371–37382. https://doi.org/10.1074/jbc.M112.367003

Stanhill A, Haynes CM, Zhang Y, Min G, Steele MC, Kalinina J, Martinez E, Pickart CM, Kong XP, Ron D (2006) An arsenite-inducible 19S regulatory particle-associated protein adapts proteasomes to proteotoxicity. Mol Cell 23(6):875–885. https://doi.org/10.1016/j.molcel.2006.07.023

Stein A, Ruggiano A, Carvalho P, Rapoport TA (2014) Key steps in ERAD of luminal ER proteins reconstituted with purified components. Cell 158(6):1375–1388. https://doi.org/10.1016/j.cell.2014.07.050

Stolz A, Hilt W, Buchberger A, Wolf DH (2011) Cdc48: a power machine in protein degradation. Trends Biochem Sci 36(10):515–523. https://doi.org/10.1016/j.tibs.2011.06.001

Sun S, Liu S, Zhang Z, Zeng W, Sun C, Tao T, Lin X, Feng XH (2017) Phosphatase UBLCP1 controls proteasome assembly. Open Biol 7(5). https://doi.org/10.1098/rsob.170042

Sun XM, Butterworth M, MacFarlane M, Dubiel W, Ciechanover A, Cohen GM (2004) Caspase activation inhibits proteasome function during apoptosis. Mol Cell 14(1):81–93. https://doi.org/10.1016/s1097-2765(04)00156-x

Tai HC, Schuman EM (2008) Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. Nat Rev Neurosci 9(11):826–838. https://doi.org/10.1038/nrn2499

Takagi K, Kim S, Yukii H, Ueno M, Morishita R, Endo Y, Kato K, Tanaka K, Saeki Y, Mizushima T (2012) Structural basis for specific recognition of Rpt1p, an ATPase subunit of 26 S proteasome, by proteasome-dedicated chaperone Hsm3p. J Biol Chem 287(15):12172–12182. https://doi.org/10.1074/jbc.M112.345876

Takeda K, Yanagida M (2005) Regulation of nuclear proteasome by Rhp6/Ubc2 through ubiquitination and destruction of the sensor and anchor Cut8. Cell 122(3):393–405. https://doi.org/10.1016/j.cell.2005.05.023

Takeuchi J, Chen H, Coffino P (2007) Proteasome substrate degradation requires association plus extended peptide. EMBO J 26(1):123–131. https://doi.org/10.1038/sj.emboj.7601476

Tanahashi N, Murakami Y, Minami Y, Shimbara N, Hendil KB, Tanaka K (2000) Hybrid proteasomes. Induction by interferon-gamma and contribution to ATP-dependent proteolysis. J Biol Chem 275(19):14336–14345. https://doi.org/10.1074/jbc.275.19.14336

Tanaka K (2009) The proteasome: overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci 85(1):12–36. https://doi.org/10.2183/pjab.85.12

Tanaka K, Ichihara A (1989) Half-life of proteasomes (multiprotease complexes) in rat liver. Biochem Biophys Res Commun 159(3):1309–1315. https://doi.org/10.1016/0006-291x(89)92253-5

Tanaka K, Waxman L, Goldberg AL (1983) ATP serves two distinct roles in protein degradation in reticulocytes, one requiring and one independent of ubiquitin. J Cell Biol 96(6):1580–1585. https://doi.org/10.1083/jcb.96.6.1580

Tanaka K, Yoshimura T, Tamura T, Fujiwara T, Kumatori A, Ichihara A (1990) Possible mechanism of nuclear translocation of proteasomes. FEBS Lett 271(1–2):41–46. https://doi.org/10.1016/0014-5793(90)80367-r

Tang WK, Li D, Li CC, Esser L, Dai R, Guo L, Xia D (2010) A novel ATP-dependent conformation in p97 N-D1 fragment revealed by crystal structures of disease-related mutants. EMBO J 29(13):2217–2229. https://doi.org/10.1038/emboj.2010.104

Tang WK, Xia D (2016) Mutations in the Human AAA(+) chaperone p97 and related diseases. Front Mol Biosci 3:79. https://doi.org/10.3389/fmolb.2016.00079

Taylor EB, Rutter J (2011) Mitochondrial quality control by the ubiquitin-proteasome system. Biochem Soc Trans 39(5):1509–1513. https://doi.org/10.1042/BST0391509

Thompson D, Hakala K, DeMartino GN (2009) Subcomplexes of PA700, the 19 S regulator of the 26 S proteasome, reveal relative roles of AAA subunits in 26 S proteasome assembly and activation and ATPase activity. J Biol Chem 284(37):24891–24903. https://doi.org/10.1074/jbc.M109.023218

Thomsen ND, Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. Cell 139(3):523–534. https://doi.org/10.1016/j.cell.2009.08.043

Thrower JS, Hoffman L, Rechsteiner M, Pickart CM (2000) Recognition of the polyubiquitin proteolytic signal. EMBO J 19(1):94–102. https://doi.org/10.1093/emboj/19.1.94

Tian G, Park S, Lee MJ, Huck B, McAllister F, Hill CP, Gygi SP, Finley D (2011) An asymmetric interface between the regulatory and core particles of the proteasome. Nat Struct Mol Biol 18(11):1259–1267. https://doi.org/10.1038/nsmb.2147

Tian L, Holmgren RA, Matouschek A (2005) A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-kappaB. Nat Struct Mol Biol 12(12):1045–1053. https://doi.org/10.1038/nsmb1018

Tian Z, Zheng H, Li J, Li Y, Su H, Wang X (2012) Genetically induced moderate inhibition of the proteasome in cardiomyocytes exacerbates myocardial ischemia-reperfusion injury in mice. Circ Res 111(5):532–542. https://doi.org/10.1161/CIRCRESAHA.112.270983

Tomaru U, Takahashi S, Ishizu A, Miyatake Y, Gohda A, Suzuki S, Ono A, Ohara J, Baba T, Murata S, Tanaka K, Kasahara M (2012) Decreased proteasomal activity causes age-related phenotypes and promotes the development of metabolic abnormalities. Am J Pathol 180(3):963–972. https://doi.org/10.1016/j.ajpath.2011.11.012

Tomita T, Hirayama S, Sakurai Y, Ohte Y, Yoshihara H, Saeki Y, Hamazaki J, Murata S (2019) Specific modification of aged proteasomes revealed by tag-exchangeable knock-in mice. Mol Cell Biol 39(1). https://doi.org/10.1128/mcb.00426-18

Tomko RJ Jr, Hochstrasser M (2011) Incorporation of the Rpn12 subunit couples completion of proteasome regulatory particle lid assembly to lid-base joining. Mol Cell 44(6):907–917. https://doi.org/10.1016/j.molcel.2011.11.020

Tomko RJ Jr, Hochstrasser M (2013) Molecular architecture and assembly of the eukaryotic proteasome. Annu Rev Biochem 82:415–445. https://doi.org/10.1146/annurev-biochem-060410-150257

Tomko RJ Jr, Hochstrasser M (2014) The intrinsically disordered Sem1 protein functions as a molecular tether during proteasome lid biogenesis. Mol Cell 53(3):433–443. https://doi.org/10.1016/j.molcel.2013.12.009

Tomko RJ Jr, Taylor DW, Chen ZA, Wang HW, Rappsilber J, Hochstrasser M (2015) A single alpha helix drives extensive remodeling of the proteasome lid and completion of regulatory particle assembly. Cell 163(2):432–444. https://doi.org/10.1016/j.cell.2015.09.022

Tonoki A, Kuranaga E, Tomioka T, Hamazaki J, Murata S, Tanaka K, Miura M (2009) Genetic evidence linking age-dependent attenuation of the 26S proteasome with the aging process. Mol Cell Biol 29(4):1095–1106. https://doi.org/10.1128/MCB.01227-08

Trader DJ, Simanski S, Kodadek T (2015) A reversible and highly selective inhibitor of the proteasomal ubiquitin receptor rpn13 is toxic to multiple myeloma cells. J Am Chem Soc 137(19):6312–6319. https://doi.org/10.1021/jacs.5b02069

Tran HJ, Allen MD, Lowe J, Bycroft M (2003) Structure of the Jab1/MPN domain and its implications for proteasome function. Biochemistry 42(39):11460–11465. https://doi.org/10.1021/bi035033g

Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell 101(3):249–258. https://doi.org/10.1016/s0092-8674(00)80835-1

Tsai FT, Hill CP (2020) Same structure, different mechanisms? Elife 9. https://doi.org/10.7554/elife.56501

Tsai YC, Weissman AM (2010) The unfolded protein response, degradation from endoplasmic reticulum and cancer. Genes Cancer 1(7):764–778. https://doi.org/10.1177/1947601910383011

Tseng BP, Green KN, Chan JL, Blurton-Jones M, LaFerla FM (2008) Abeta inhibits the proteasome and enhances amyloid and tau accumulation. Neurobiol Aging 29(11):1607–1618. https://doi.org/10.1016/j.neurobiolaging.2007.04.014

Tsuchiya H, Ohtake F, Arai N, Kaiho A, Yasuda S, Tanaka K, Saeki Y (2017) In Vivo ubiquitin linkage-type analysis reveals that the Cdc48-Rad23/Dsk2 axis contributes to K48-linked chain specificity of the proteasome. Mol Cell 66(4):488–502 e487. https://doi.org/10.1016/j.molcel.2017.04.024

Turakhiya A, Meyer SR, Marincola G, Bohm S, Vanselow JT, Schlosser A, Hofmann K, Buchberger A (2018) ZFAND1 recruits p 97 and the 26S proteasome to promote the clearance of arsenite-induced stress granules. Mol Cell 70(5):906–919 e907. https://doi.org/10.1016/j.molcel.2018.04.021

Twomey EC, Ji Z, Wales TE, Bodnar NO, Ficarro SB, Marto JA, Engen JR, Rapoport TA (2019) Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. Science 365(6452). https://doi.org/10.1126/science.aax1033

Uddin MM, Ohigashi I, Motosugi R, Nakayama T, Sakata M, Hamazaki J, Nishito Y, Rode I, Tanaka K, Takemoto T, Murata S, Takahama Y (2017) Foxn1-beta5t transcriptional axis controls CD8(+) T-cell production in the thymus. Nat Commun 8:14419. https://doi.org/10.1038/ncomms14419

Uechi H, Hamazaki J, Murata S (2014) Characterization of the testis-specific proteasome subunit alpha4s in mammals. J Biol Chem 289(18):12365–12374. https://doi.org/10.1074/jbc.M114.558866

Ungvari Z, Csiszar A, Sosnowska D, Philipp EE, Campbell CM, McQuary PR, Chow TT, Coelho M, Didier ES, Gelino S, Holmbeck MA, Kim I, Levy E, Sonntag WE, Whitby PW, Austad SN, Ridgway I (2013) Testing predictions of the oxidative stress hypothesis of aging using a novel invertebrate model of longevity: the giant clam (Tridacna derasa). J Gerontol A Biol Sci Med Sci 68(4):359–367. https://doi.org/10.1093/gerona/gls159

Unno M, Mizushima T, Morimoto Y, Tomisugi Y, Tanaka K, Yasuoka N, Tsukihara T (2002) The structure of the mammalian 20S proteasome at 2.75 A resolution. Structure 10(5):609–618

Unverdorben P, Beck F, Sledz P, Schweitzer A, Pfeifer G, Plitzko JM, Baumeister W, Forster F (2014) Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. Proc Natl Acad Sci U S A 111(15):5544–5549. https://doi.org/10.1073/pnas.1403409111

Ustun S, Hafren A, Liu Q, Marshall RS, Minina EA, Bozhkov PV, Vierstra RD, Hofius D (2018) Bacteria exploit autophagy for proteasome degradation and enhanced virulence in plants. Plant Cell 30(3):668–685. https://doi.org/10.1105/tpc.17.00815

Vale RD (2000) AAA proteins. Lords of the ring. J Cell Biol 150(1):F13–19. https://doi.org/10.1083/jcb.150.1.f13

van den Boom J, Meyer H (2018) VCP/p97-mediated unfolding as a principle in protein homeostasis and signaling. Mol Cell 69(2):182–194. https://doi.org/10.1016/j.molcel.2017.10.028

van der Lee R, Lang B, Kruse K, Gsponer J, Sanchez de Groot N, Huynen MA, Matouschek A, Fuxreiter M, Babu MM (2014) Intrinsically disordered segments affect protein half-life in the cell and during evolution. Cell Rep 8(6):1832–1844. https://doi.org/10.1016/j.celrep.2014.07.055

van Deventer S, Menendez-Benito V, van Leeuwen F, Neefjes J (2015) N-terminal acetylation and replicative age affect proteasome localization and cell fitness during aging. J Cell Sci 128(1):109–117. https://doi.org/10.1242/jcs.157354

van Nocker S, Sadis S, Rubin DM, Glickman M, Fu H, Coux O, Wefes I, Finley D, Vierstra RD (1996) The multiubiquitin-chain-binding protein Mcb1 is a component of the 26S proteasome in Saccharomyces cerevisiae and plays a nonessential, substrate-specific role in protein turnover. Mol Cell Biol 16(11):6020–6028. https://doi.org/10.1128/mcb.16.11.6020

Vander Linden RT, Hemmis CW, Schmitt B, Ndoja A, Whitby FG, Robinson H, Cohen RE, Yao T, Hill CP (2015) Structural basis for the activation and inhibition of the UCH37 deubiquitylase. Mol Cell 57(5):901–911. https://doi.org/10.1016/j.molcel.2015.01.016

Vangala JR, Dudem S, Jain N, Kalivendi SV (2014) Regulation of PSMB5 protein and beta subunits of mammalian proteasome by constitutively activated signal transducer and activator of transcription 3 (STAT3): potential role in bortezomib-mediated anticancer therapy. J Biol Chem 289(18):12612–12622. https://doi.org/10.1074/jbc.M113.542829

Varadan R, Assfalg M, Haririnia A, Raasi S, Pickart C, Fushman D (2004) Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. J Biol Chem 279(8):7055–7063. https://doi.org/10.1074/jbc.M309184200

Varadan R, Assfalg M, Raasi S, Pickart C, Fushman D (2005) Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. Mol Cell 18(6):687–698. https://doi.org/10.1016/j.molcel.2005.05.013

Vashist S, Ng DT (2004) Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control. J Cell Biol 165(1):41–52. https://doi.org/10.1083/jcb.200309132

Vekaria PH, Home T, Weir S, Schoenen FJ, Rao R (2016) Targeting p97 to disrupt protein homeostasis in cancer. Front Oncol 6:181. https://doi.org/10.3389/fonc.2016.00181

Velichutina I, Connerly PL, Arendt CS, Li X, Hochstrasser M (2004) Plasticity in eucaryotic 20S proteasome ring assembly revealed by a subunit deletion in yeast. EMBO J 23(3):500–510. https://doi.org/10.1038/sj.emboj.7600059

Verhoef LG, Heinen C, Selivanova A, Halff EF, Salomons FA, Dantuma NP (2009) Minimal length requirement for proteasomal degradation of ubiquitin-dependent substrates. FASEB J 23(1):123–133. https://doi.org/10.1096/fj.08-115055

Verma R, Aravind L, Oania R, McDonald WH, Yates JR 3rd, Koonin EV, Deshaies RJ (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. Science 298(5593):611–615. https://doi.org/10.1126/science.1075898

Verma R, Chen S, Feldman R, Schieltz D, Yates J, Dohmen J, Deshaies RJ (2000) Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. Mol Biol Cell 11(10):3425–3439. https://doi.org/10.1091/mbc.11.10.3425

Verma R, Oania R, Graumann J, Deshaies RJ (2004) Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. Cell 118(1):99–110. https://doi.org/10.1016/j.cell.2004.06.014

Verma R, Oania RS, Kolawa NJ, Deshaies RJ (2013) Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. Elife 2:e00308. https://doi.org/10.7554/eLife.00308

Vernace VA, Arnaud L, Schmidt-Glenewinkel T, Figueiredo-Pereira ME (2007) Aging perturbs 26S proteasome assembly in Drosophila melanogaster. FASEB J 21(11):2672–2682. https://doi.org/10.1096/fj.06-6751com

VerPlank JJS, Goldberg AL (2017) Regulating protein breakdown through proteasome phosphorylation. Biochem J 474(19):3355–3371. https://doi.org/10.1042/BCJ20160809

VerPlank JJS, Lokireddy S, Zhao J, Goldberg AL (2019) 26S Proteasomes are rapidly activated by diverse hormones and physiological states that raise cAMP and cause Rpn6 phosphorylation. Proc Natl Acad Sci U S A 116(10):4228–4237. https://doi.org/10.1073/pnas.1809254116

Vilchez D, Boyer L, Morantte I, Lutz M, Merkwirth C, Joyce D, Spencer B, Page L, Masliah E, Berggren WT, Gage FH, Dillin A (2012) Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. Nature 489(7415):304–308. https://doi.org/10.1038/nature11468

Vilchez D, Saez I, Dillin A (2014) The role of protein clearance mechanisms in organismal ageing and age-related diseases. Nat Commun 5:5659. https://doi.org/10.1038/ncomms6659

Vinitsky A, Cardozo C, Sepp-Lorenzino L, Michaud C, Orlowski M (1994) Inhibition of the proteolytic activity of the multicatalytic proteinase complex (proteasome) by substrate-related peptidyl aldehydes. J Biol Chem 269(47):29860–29866

Vinitsky A, Michaud C, Powers JC, Orlowski M (1992) Inhibition of the chymotrypsin-like activity of the pituitary multicatalytic proteinase complex. Biochemistry 31(39):9421–9428. https://doi.org/10.1021/bi00154a014

Voges D, Zwickl P, Baumeister W (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. Annu Rev Biochem 68:1015–1068. https://doi.org/10.1146/annurev.biochem.68.1.1015

Voorhees PM, Orlowski RZ (2006) The proteasome and proteasome inhibitors in cancer therapy. Annu Rev Pharmacol Toxicol 46:189–213. https://doi.org/10.1146/annurev.pharmtox.46.120604.141300

Voutsadakis IA (2017) Proteasome expression and activity in cancer and cancer stem cells. Tumour Biol 39(3):1010428317692248. https://doi.org/10.1177/1010428317692248

Waite KA, De-La Mota-Peynado A, Vontz G, Roelofs J (2016) Starvation induces proteasome autophagy with different pathways for core and regulatory particles. J Biol Chem 291(7):3239–3253. https://doi.org/10.1074/jbc.M115.699124

Walters KJ, Kleijnen MF, Goh AM, Wagner G, Howley PM (2002) Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. Biochemistry 41(6):1767–1777. https://doi.org/10.1021/bi011892y

Walters KJ, Lech PJ, Goh AM, Wang Q, Howley PM (2003) DNA-repair protein hHR23a alters its protein structure upon binding proteasomal subunit S5a. Proc Natl Acad Sci U S A 100(22):12694–12699. https://doi.org/10.1073/pnas.1634989100

Wang CY, Mayo MW, Baldwin AS Jr (1996) TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF-kappaB. Science 274(5288):784–787. https://doi.org/10.1126/science.274.5288.784

Wang CY, Mayo MW, Korneluk RG, Goeddel DV, Baldwin AS Jr (1998) NF-kappaB antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. Science 281(5383):1680–1683. https://doi.org/10.1126/science.281.5383.1680

Wang H, Lim PJ, Yin C, Rieckher M, Vogel BE, Monteiro MJ (2006) Suppression of polyglutamine-induced toxicity in cell and animal models of Huntington's disease by ubiquilin. Hum Mol Genet 15(6):1025–1041. https://doi.org/10.1093/hmg/ddl017

Wang H, Monteiro MJ (2007) Ubiquilin interacts and enhances the degradation of expanded-polyglutamine proteins. Biochem Biophys Res Commun 360(2):423–427. https://doi.org/10.1016/j.bbrc.2007.06.097

Wang J, Hartling JA, Flanagan JM (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. Cell 91 (4):447–456. https://doi.org/10.1016/s0092-8674(00)80431-6

Wang WL, Yu Z, Castillo-Menendez LR, Sodroski J, Mao Y (2019) Robustness of signal detection in cryo-electron microscopy via a bi-objective-function approach. BMC Bioinformatics 20(1):169. https://doi.org/10.1186/s12859-019-2714-8

Wang X, Chen CF, Baker PR, Chen PL, Kaiser P, Huang L (2007) Mass spectrometric characterization of the affinity-purified human 26S proteasome complex. Biochemistry 46(11):3553–3565. https://doi.org/10.1021/bi061994u

Wang X, Xu H, Ju D, Xie Y (2008) Disruption of Rpn4-induced proteasome expression in Saccharomyces cerevisiae reduces cell viability under stressed conditions. Genetics 180(4):1945–1953. https://doi.org/10.1534/genetics.108.094524

Wang X, Yen J, Kaiser P, Huang L (2010) Regulation of the 26S proteasome complex during oxidative stress. Sci Signal 3(151):ra88. https://doi.org/10.1126/scisignal.2001232

Wang Z, Yang J, Kirk C, Fang Y, Alsina M, Badros A, Papadopoulos K, Wong A, Woo T, Bomba D, Li J, Infante JR (2013) Clinical pharmacokinetics, metabolism, and drug-drug interaction of carfilzomib. Drug Metab Dispos 41(1):230–237. https://doi.org/10.1124/dmd.112.047662

Wani PS, Rowland MA, Ondracek A, Deeds EJ, Roelofs J (2015) Maturation of the proteasome core particle induces an affinity switch that controls regulatory particle association. Nat Commun 6:6384. https://doi.org/10.1038/ncomms7384

Wani PS, Suppahia A, Capalla X, Ondracek A, Roelofs J (2016) Phosphorylation of the C-terminal tail of proteasome subunit alpha7 is required for binding of the proteasome quality control factor Ecm29. Sci Rep 6:27873. https://doi.org/10.1038/srep27873

Webb AE, Brunet A (2014) FOXO transcription factors: key regulators of cellular quality control. Trends Biochem Sci 39(4):159–169. https://doi.org/10.1016/j.tibs.2014.02.003

Weberruss MH, Savulescu AF, Jando J, Bissinger T, Harel A, Glickman MH, Enenkel C (2013) Blm10 facilitates nuclear import of proteasome core particles. EMBO J 32(20):2697–2707. https://doi.org/10.1038/emboj.2013.192

Wehmer M, Rudack T, Beck F, Aufderheide A, Pfeifer G, Plitzko JM, Forster F, Schulten K, Baumeister W, Sakata E (2017) Structural insights into the functional cycle of the ATPase module of the 26S proteasome. Proc Natl Acad Sci U S A 114(6):1305–1310. https://doi.org/10.1073/pnas.1621129114

Wendler P, Ciniawsky S, Kock M, Kube S (2012) Structure and function of the AAA + nucleotide binding pocket. Biochim Biophys Acta 1823(1):2–14. 10.1016/j.bbamcr.2011.06.014

Wendler P, Lehmann A, Janek K, Baumgart S, Enenkel C (2004) The bipartite nuclear localization sequence of Rpn2 is required for nuclear import of proteasomal base complexes via karyopherin alphabeta and proteasome functions. J Biol Chem 279(36):37751–37762. https://doi.org/10.1074/jbc.M403551200

Wenzel T, Eckerskorn C, Lottspeich F, Baumeister W (1994) Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. FEBS Lett 349(2):205–209. https://doi.org/10.1016/0014-5793(94)00665-2

Whitby FG, Masters EI, Kramer L, Knowlton JR, Yao Y, Wang CC, Hill CP (2000) Structural basis for the activation of 20S proteasomes by 11S regulators. Nature 408(6808):115–120. https://doi.org/10.1038/35040607

Wilk S, Orlowski M (1983) Evidence that pituitary cation-sensitive neutral endopeptidase is a multicatalytic protease complex. J Neurochem 40(3):842–849. https://doi.org/10.1111/j.1471-4159.1983.tb08056.x

Wilkinson KD, Urban MK, Haas AL (1980) Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. J Biol Chem 255(16):7529–7532

Williamson A, Wickliffe KE, Mellone BG, Song L, Karpen GH, Rape M (2009) Identification of a physiological E2 module for the human anaphase-promoting complex. Proc Natl Acad Sci U S A 106(43):18213–18218. https://doi.org/10.1073/pnas.0907887106

Winston JT, Strack P, Beer-Romero P, Chu CY, Elledge SJ, Harper JW (1999) The SCFbeta-TRCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in IkappaBalpha and beta-catenin and stimulates IkappaBalpha ubiquitination in vitro. Genes Dev 13(3):270–283. https://doi.org/10.1101/gad.13.3.270

Witt E, Zantopf D, Schmidt M, Kraft R, Kloetzel PM, Kruger E (2000) Characterisation of the newly identified human Ump1 homologue POMP and analysis of LMP7(beta 5i) incorporation into 20 S proteasomes. J Mol Biol 301(1):1–9. https://doi.org/10.1006/jmbi.2000.3959

Witt S, Kwon YD, Sharon M, Felderer K, Beuttler M, Robinson CV, Baumeister W, Jap BK (2006) Proteasome assembly triggers a switch required for active-site maturation. Structure 14(7):1179–1188. https://doi.org/10.1016/j.str.2006.05.019

Woelk T, Oldrini B, Maspero E, Confalonieri S, Cavallaro E, Di Fiore PP, Polo S (2006) Molecular mechanisms of coupled monoubiquitination. Nat Cell Biol 8(11):1246–1254. https://doi.org/10.1038/ncb1484

Wolf DH, Stolz A (2012) The Cdc48 machine in endoplasmic reticulum associated protein degradation. Biochim Biophys Acta 1823(1):117–124. https://doi.org/10.1016/j.bbamcr.2011.09.002

Worden EJ, Dong KC, Martin A (2017) An AAA motor-driven mechanical switch in Rpn11 controls deubiquitination at the 26S proteasome. Mol Cell 67(5):799–811 e798. https://doi.org/10.1016/j.molcel.2017.07.023

Worden EJ, Padovani C, Martin A (2014) Structure of the Rpn11-Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. Nat Struct Mol Biol 21(3):220–227. https://doi.org/10.1038/nsmb.2771

Wright KL, White LC, Kelly A, Beck S, Trowsdale J, Ting JP (1995) Coordinate regulation of the human TAP1 and LMP2 genes from a shared bidirectional promoter. J Exp Med 181(4):1459–1471. https://doi.org/10.1084/jem.181.4.1459

Wu J, Ma YB, Congdon C, Brett B, Chen S, Xu Y, Ouyang Q, Mao Y (2017) Massively parallel unsupervised single-particle cryo-EM data clustering via statistical manifold learning. PLoS ONE 12(8):e0182130. https://doi.org/10.1371/journal.pone.0182130

Wu T, Merbl Y, Huo Y, Gallop JL, Tzur A, Kirschner MW (2010) UBE2S drives elongation of K11-linked ubiquitin chains by the anaphase-promoting complex. Proc Natl Acad Sci U S A 107(4):1355–1360. https://doi.org/10.1073/pnas.0912802107

Wu X, Karin M (2015) Emerging roles of Lys63-linked polyubiquitylation in immune responses. Immunol Rev 266(1):161–174. https://doi.org/10.1111/imr.12310

Wu X, Rapoport TA (2018) Mechanistic insights into ER-associated protein degradation. Curr Opin Cell Biol 53:22–28. https://doi.org/10.1016/j.ceb.2018.04.004

Wu Y, Luo H, Kanaan N, Wu J (2000) The proteasome controls the expression of a proliferation-associated nuclear antigen Ki-67. J Cell Biochem 76(4):596–604

Xia D, Tang WK, Ye Y (2016) Structure and function of the AAA + ATPase p97/Cdc48p. Gene 583(1):64–77. https://doi.org/10.1016/j.gene.2016.02.042

Xie Y, Varshavsky A (2000) Physical association of ubiquitin ligases and the 26S proteasome. Proc Natl Acad Sci U S A 97(6):2497–2502. https://doi.org/10.1073/pnas.060025497

Xie Y, Varshavsky A (2001) RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit. Proc Natl Acad Sci U S A 98(6):3056–3061. https://doi.org/10.1073/pnas.071022298

Xie Y, Varshavsky A (2002) UFD4 lacking the proteasome-binding region catalyses ubiquitination but is impaired in proteolysis. Nat Cell Biol 4(12):1003–1007. https://doi.org/10.1038/ncb889

Xu H, Fu J, Ha SW, Ju D, Zheng J, Li L, Xie Y (2012) The CCAAT box-binding transcription factor NF-Y regulates basal expression of human proteasome genes. Biochim Biophys Acta 1823(4):818–825. https://doi.org/10.1016/j.bbamcr.2012.01.002

Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, Rush J, Hochstrasser M, Finley D, Peng J (2009) Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. Cell 137(1):133–145. https://doi.org/10.1016/j.cell.2009.01.041

Xu S, Peng G, Wang Y, Fang S, Karbowski M (2011) The AAA-ATPase p97 is essential for outer mitochondrial membrane protein turnover. Mol Biol Cell 22(3):291–300. https://doi.org/10.1091/mbc.E10-09-0748

Xu Y, Wu J, Yin CC, Mao Y (2016) Unsupervised Cryo-EM data clustering through adaptively constrained K-means algorithm. PLoS ONE 11(12):e0167765. https://doi.org/10.1371/journal.pone.0167765

Yao T, Cohen RE (2002) A cryptic protease couples deubiquitination and degradation by the proteasome. Nature 419(6905):403–407. https://doi.org/10.1038/nature01071

Yao T, Song L, Jin J, Cai Y, Takahashi H, Swanson SK, Washburn MP, Florens L, Conaway RC, Cohen RE, Conaway JW (2008) Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the proteasome and in the Ino80 chromatin-remodeling complex. Mol Cell 31(6):909–917. https://doi.org/10.1016/j.molcel.2008.08.027

Yao T, Song L, Xu W, DeMartino GN, Florens L, Swanson SK, Washburn MP, Conaway RC, Conaway JW, Cohen RE (2006) Proteasome recruitment and activation of the Uch37

deubiquitinating enzyme by Adrm1. Nat Cell Biol 8(9):994–1002. https://doi.org/10.1038/ncb1460

Yashiroda H, Mizushima T, Okamoto K, Kameyama T, Hayashi H, Kishimoto T, Niwa S, Kasahara M, Kurimoto E, Sakata E, Takagi K, Suzuki A, Hirano Y, Murata S, Kato K, Yamane T, Tanaka K (2008) Crystal structure of a chaperone complex that contributes to the assembly of yeast 20S proteasomes. Nat Struct Mol Biol 15(3):228–236. https://doi.org/10.1038/nsmb.1386

Yasuda S, Tsuchiya H, Kaiho A, Guo Q, Ikeuchi K, Endo A, Arai N, Ohtake F, Murata S, Inada T, Baumeister W, Fernandez-Busnadiego R, Tanaka K, Saeki Y (2020) Stress- and ubiquitylation-dependent phase separation of the proteasome. Nature 578(7794):296–300. https://doi.org/10.1038/s41586-020-1982-9

Yau R, Rape M (2016) The increasing complexity of the ubiquitin code. Nat Cell Biol 18(6):579–586. https://doi.org/10.1038/ncb3358

Yau RG, Doerner K, Castellanos ER, Haakonsen DL, Werner A, Wang N, Yang XW, Martinez-Martin N, Matsumoto ML, Dixit VM, Rape M (2017) Assembly and function of heterotypic ubiquitin chains in cell-cycle and protein quality control. Cell 171(4):918–933 e920. https://doi.org/10.1016/j.cell.2017.09.040

Ye Y, Rape M (2009) Building ubiquitin chains: E2 enzymes at work. Nat Rev Mol Cell Biol 10(11):755–764. https://doi.org/10.1038/nrm2780

Yedidi RS, Fatehi AK, Enenkel C (2016) Proteasome dynamics between proliferation and quiescence stages of Saccharomyces cerevisiae. Crit Rev Biochem Mol Biol 51(6):497–512. https://doi.org/10.1080/10409238.2016.1230087

Yokoi M, Hanaoka F (2017) Two mammalian homologs of yeast Rad23, HR23A and HR23B, as multifunctional proteins. Gene 597:1–9. https://doi.org/10.1016/j.gene.2016.10.027

Yu H, Kago G, Yellman CM, Matouschek A (2016a) Ubiquitin-like domains can target to the proteasome but proteolysis requires a disordered region. EMBO J 35(14):1522–1536. https://doi.org/10.15252/embj.201593147

Yu H, Lupoli TJ, Kovach A, Meng X, Zhao G, Nathan CF, Li H (2018) ATP hydrolysis-coupled peptide translocation mechanism of Mycobacterium tuberculosis ClpB. Proc Natl Acad Sci U S A 115(41):E9560–E9569. https://doi.org/10.1073/pnas.1810648115

Yu H, Matouschek A (2017) Recognition of Client Proteins by the Proteasome. Annu Rev Biophys 46:149–173. https://doi.org/10.1146/annurev-biophys-070816-033719

Yu H, Singh Gautam AK, Wilmington SR, Wylie D, Martinez-Fonts K, Kago G, Warburton M, Chavali S, Inobe T, Finkelstein IJ, Babu MM, Matouschek A (2016b) Conserved Sequence Preferences Contribute to Substrate Recognition by the Proteasome. J Biol Chem 291(28):14526–14539. https://doi.org/10.1074/jbc.M116.727578

Yu Y, Smith DM, Kim HM, Rodriguez V, Goldberg AL, Cheng Y (2010) Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome-ATPase interactions. EMBO J 29(3):692–702. https://doi.org/10.1038/emboj.2009.382

Yun C, Stanhill A, Yang Y, Zhang Y, Haynes CM, Xu CF, Neubert TA, Mor A, Philips MR, Ron D (2008) Proteasomal adaptation to environmental stress links resistance to proteotoxicity with longevity in Caenorhabditis elegans. Proc Natl Acad Sci U S A 105(19):7094–7099. https://doi.org/10.1073/pnas.0707025105

Yu Z, Yu Y, Wang F, Myasnikov AG, Coffino P, Cheng Y (2020) Allosteric coupling between alpha-rings of the 20S proteasome. Nat Commun 11(1):4580. https://doi.org/10.1038/s41467-020-18415-7

Zaiss DM, Standera S, Holzhutter H, Kloetzel P, Sijts AJ (1999) The proteasome inhibitor PI31 competes with PA28 for binding to 20S proteasomes. FEBS Lett 457(3):333–338. https://doi.org/10.1016/s0014-5793(99)01072-8

Zaiss DM, Standera S, Kloetzel PM, Sijts AJ (2002) PI31 is a modulator of proteasome formation and antigen processing. Proc Natl Acad Sci U S A 99(22):14344–14349. https://doi.org/10.1073/pnas.212257299

Zanelli E, Zhou P, Cao H, Smart MK, David CS (1993) Genomic organization and tissue expression of the mouse proteasome gene Lmp-7. Immunogenetics 38(6):400–407. https://doi.org/10.1007/bf00184520

Zhang D, Chen T, Ziv I, Rosenzweig R, Matiuhin Y, Bronner V, Glickman MH, Fushman D (2009a) Together, Rpn10 and Dsk2 can serve as a polyubiquitin chain-length sensor. Mol Cell 36(6):1018–1033. https://doi.org/10.1016/j.molcel.2009.11.012

Zhang F, Hu M, Tian G, Zhang P, Finley D, Jeffrey PD, Shi Y (2009b) Structural insights into the regulatory particle of the proteasome from Methanocaldococcus jannaschii. Mol Cell 34(4):473–484. https://doi.org/10.1016/j.molcel.2009.04.021

Zhang F, Su K, Yang X, Bowe DB, Paterson AJ, Kudlow JE (2003) O-GlcNAc modification is an endogenous inhibitor of the proteasome. Cell 115(6):715–725. https://doi.org/10.1016/s0092-8674(03)00974-7

Zhang F, Wu Z, Zhang P, Tian G, Finley D, Shi Y (2009c) Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from Methanocaldococcus jannaschii. Mol Cell 34(4):485–496. https://doi.org/10.1016/j.molcel.2009.04.022

Zhang L, Chen S, Ruan J, Wu J, Tong AB, Yin Q, Li Y, David L, Lu A, Wang WL, Marks C, Ouyang Q, Zhang X, Mao Y, Wu H (2015) Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. Science 350(6259):404–409. https://doi.org/10.1126/science.aac5789

Zhang N, Wang Q, Ehlinger A, Randles L, Lary JW, Kang Y, Haririnia A, Storaska AJ, Cole JL, Fushman D, Walters KJ (2009d) Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. Mol Cell 35(3):280–290. https://doi.org/10.1016/j.molcel.2009.06.010

Zhang NY, Jacobson AD, Macfadden A, Liu CW (2011) Ubiquitin chain trimming recycles the substrate binding sites of the 26 S proteasome and promotes degradation of lysine 48-linked polyubiquitin conjugates. J Biol Chem 286(29):25540–25546. https://doi.org/10.1074/jbc.M111.260505

Zhang S, Mao Y (2020) AAA + ATPases in Protein Degradation: Structures, Functions and Mechanisms. Biomolecules 10 (4). https://doi.org/10.3390/biom10040629

Zhang X, Shaw A, Bates PA, Newman RH, Gowen B, Orlova E, Gorman MA, Kondo H, Dokurno P, Lally J, Leonard G, Meyer H, van Heel M, Freemont PS (2000) Structure of the AAA ATPase p97. Mol Cell 6(6):1473–1484

Zhang X, Wigley DB (2008) The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA + proteins. Nat Struct Mol Biol 15(11):1223–1227. https://doi.org/10.1038/nsmb.1501

Zhang Y, Manning BD (2015) mTORC1 signaling activates NRF1 to increase cellular proteasome levels. Cell Cycle 14(13):2011–2017. https://doi.org/10.1080/15384101.2015.1044188

Zhang Y, Nicholatos J, Dreier JR, Ricoult SJ, Widenmaier SB, Hotamisligil GS, Kwiatkowski DJ, Manning BD (2014) Coordinated regulation of protein synthesis and degradation by mTORC1. Nature 513(7518):440–443. https://doi.org/10.1038/nature13492

Zhang Z, Lv X, Yin WC, Zhang X, Feng J, Wu W, Hui CC, Zhang L, Zhao Y (2013) Ter94 ATPase complex targets k11-linked ubiquitinated ci to proteasomes for partial degradation. Dev Cell 25(6):636–644. https://doi.org/10.1016/j.devcel.2013.05.006

Zhao J, Zhai B, Gygi SP, Goldberg AL (2015) mTOR inhibition activates overall protein degradation by the ubiquitin proteasome system as well as by autophagy. Proc Natl Acad Sci U S A 112(52):15790–15797. https://doi.org/10.1073/pnas.1521919112

Zhao M, Zhang NY, Zurawel A, Hansen KC, Liu CW (2010) Degradation of some polyubiquitinated proteins requires an intrinsic proteasomal binding element in the substrates. J Biol Chem 285(7):4771–4780. https://doi.org/10.1074/jbc.M109.060095

Zhao S, Ulrich HD (2010) Distinct consequences of posttranslational modification by linear versus K63-linked polyubiquitin chains. Proc Natl Acad Sci U S A 107(17):7704–7709. https://doi.org/10.1073/pnas.0908764107

Zheng N, Shabek N (2017) Ubiquitin Ligases: Structure, Function, and Regulation. Annu Rev Biochem 86:129–157. https://doi.org/10.1146/annurev-biochem-060815-014922

Zhu Y, Ouyang Q, Mao Y (2017) A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. BMC Bioinformatics 18(1):348. https://doi.org/10.1186/s12859-017-1757-y

Zhu Y, Wang WL, Yu D, Ouyang Q, Lu Y, Mao Y (2018) Structural mechanism for nucleotide-driven remodeling of the AAA-ATPase unfoldase in the activated human 26S proteasome. Nat Commun 9(1):1360. https://doi.org/10.1038/s41467-018-03785-w

Zong N, Ping P, Lau E, Choi HJ, Ng DC, Meyer D, Fang C, Li H, Wang D, Zelaya IM, Yates JR 3rd, Lam MP (2014) Lysine ubiquitination and acetylation of human cardiac 20S proteasomes. Proteomics Clin Appl 8(7–8):590–594. https://doi.org/10.1002/prca.201400029

Zong WX, Edelstein LC, Chen C, Bash J, Gelinas C (1999) The prosurvival Bcl-2 homolog Bfl-1/A1 is a direct transcriptional target of NF-kappaB that blocks TNFalpha-induced apoptosis. Genes Dev 13(4):382–387. https://doi.org/10.1101/gad.13.4.382

Zuhl F, Seemuller E, Golbik R, Baumeister W (1997) Dissecting the assembly pathway of the 20S proteasome. FEBS Lett 418(1–2):189–194. https://doi.org/10.1016/s0014-5793(97)01370-7

Zuin A, Bichmann A, Isasa M, Puig-Sarries P, Diaz LM, Crosas B (2015) Rpn10 monoubiquitination orchestrates the association of the ubiquilin-type DSK2 receptor with the proteasome. Biochem J 472(3):353–365. https://doi.org/10.1042/BJ20150609

Chapter 2
Factor VIII and Factor V Membrane Bound Complexes

Svetla Stoilova-McPhie

Abstract The formation of membrane-bound complexes between specific coagulation factors at different cell surfaces is required for effective blood clotting. The most important of these complexes, the intrinsic Tenase and Prothrombinase complexes, are formed on the activated platelet surface during the propagation phase of coagulation. These two complexes are highly specific in their assembly mechanism and function modulated by anionic membranes, thus offering desirable targets for pharmaceutical interventions. Factor V (FV) and factor VIII (FVIII) are highly homologous non-enzymatic proteins. In their active state, FVa and FVIIIa serve as cofactors for the respective serine proteases factor Xa (FXa) and factor IXa (FIXa), significantly increasing their catalytic activity. This is achieved by forming well organized membrane-bound complexes at the phosphatidylserine rich activated platelet membrane in the presence of Ca^{2+} ions. The tenase (FVIIIa/FIXa) complex, catalyzes the proteolytic conversion of FX to FXa. Subsequently the prothrombinase (FVa/FXa) complex catalyzes the conversion of prothrombin to thrombin, required for efficient blood clotting. Although significant knowledge of FV and FVIII biochemistry and regulation has been achieved, the molecular mechanisms of their function are yet to be defined. Understanding the geometric assembly of the tenase and prothrombinase complexes is paramount in defining the structural basis of bleeding and thrombotic disorders. Such knowledge will enable the design of efficient pro- and anticoagulant therapies critical for regulating abnormal hemostasis. In this chapter, we will summarize the findings to date, showing our achievement in the field and outlining the future findings required to grasp the complexity of these proteins.

Keywords Blood coagulation · Factor VIII · Factor V · Activated platelets · Membrane-Associated proteins · 3D structure · Cryo-Electron microscopy · Macromolecular complexes

S. Stoilova-McPhie (✉)
Center for Nanoscale Systems (CNS), Laboratory For Integrated Sciences and Engineering (LISE), Faculty of Art and Sciences (FAS), Harvard University, 11 Oxford Street, Cambridge, MA 02138, England, UK
e-mail: stoilovamcphie@fas.harvard.edu

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_2

Introduction

Haemostasis is defined as arrest of bleeding, derived from the Greek words 'haeme' meaning blood and 'stasis' meaning to stop. It is the complex physiological process responsible for stopping blood loss that depends on a delicate equilibrium between pro- and anticoagulant forces. Hemostasis' main components are cells (platelets, and fibroblasts), soluble proteins (coagulation factors) and insoluble proteins (extracellular matrix proteins). The hemostatic system is activated upon vascular injury. Primary hemostasis leads to the formation of a platelet plug at the site of the injury. Secondary hemostasis leads to the generation of a fibrin clot that allows time for the injured endothelium to heal. The final stage—fibrinolysis activates the anticoagulant mechanisms necessary to dissolve the clot and restore blood flow. Secondary hemostasis is generally referred as blood coagulation, defined as the formation of insoluble, cross-linked fibrin by a sequential activation of coagulation factors and thrombin. Fibrin also stabilizes the primary platelet plug in larger blood vessels to stop hemorrhage (Gale 2011; Palta et al. 2014).

The blood coagulation process requires the following elements:

- Cells
- Enzymatic coagulation factors (zymogenes)
- Non-enzymatic coagulation factors (pro-cofactors)
- Calcium ions (Ca^{2+})
- Phosphatidylserine (PS) rich phospholipid (PL) membrane surface.

The coagulation process was initially described as a series of sequential proteolytic reactions, in which the activation of one clotting factor leads to the activation of the next, culminating in a thrombin 'burst' and fibrin generation. This model, known as a clotting cascade or 'waterfall model' explains well the clinical tests designed to characterize the main bleeding and thrombotic disorders at the time (Macfarlane 1964; Davie and Ratnoff 1964).

Traditionally the clotting cascade has been divided into intrinsic, extrinsic and common pathways (Fig. 2.1). This division has been helpful for interpreting laboratory results. The two clinical clotting tests: Activated Partial Thromboplastin Time (APTT) (Langdellet al. 1953; Quick 1950) for the intrinsic/common pathways and the Prothrombin Time (PT) (Quick 1950) for the extrinsic/common pathways represent well the property of blood to follow different mechanism of clotting in vivo and in vitro.

The intrinsic pathway follows the coagulation of blood in vitro that occurs when blood is in contact with a foreign body like glass, silica, kaolin. The extrinsic or FVII pathway follows the in vivo coagulation, where the 'intrinsic' tenase (FVIIa/TF) complex is formed during vascular injury producing small amount of FXa and thrombin that activates the cofactors VIII and V required for the assembly of the 'extrinsic' tenase and prothrombinase complexes on the activated platelet surface.

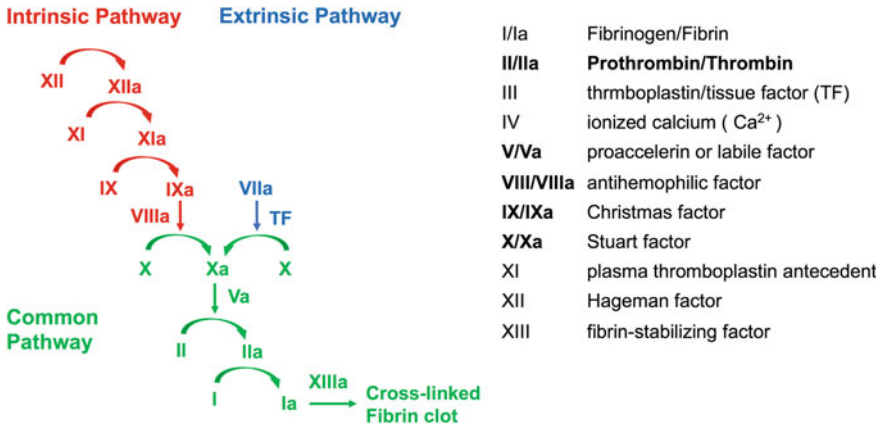


Fig. 2.1 Cascade model of blood coagulation. The coagulation factors are presented with roman numerals. The active form is denoted with a suffix 'a'. The factors in bold are the components of the tenase and prothrombinase complexes. Only the pro-coagulant pathways are indicated for clarity

These two complexes work in synchronization to generate a thrombin 'burst'—several orders of magnitude acceleration of thrombin activation, necessary for efficient hemostasis. The nomenclature of the coagulation factors participating in the clotting is shown in Fig. 2.1. In bold are the ones that participate in the 'extrinsic' tenase (FVIIIa/FIXa) and prothrombinase (FVa/FXa) complexes securing timely clotting. The significance of Ca ions (Ca^{2+}) and phospholipid (PL) surfaces for blood clotting are illustrated in Table 2.1.

Table 2.1 Clotting times of blood in vitro, in the presence and absence of different components. APPT and PT refer to the clotting tests: activated partial thromboplastin time and prothrombin time, and their components

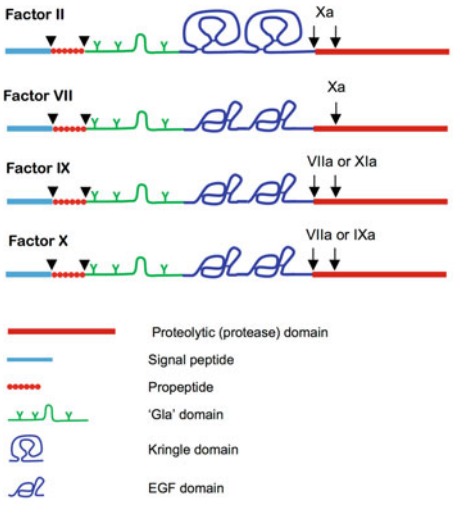
Clotting system	Clotting time
Whole blood	240–460 s (4–8 min)
Whole blood + EDTA or citrate	Infinite
Citrated platelet-poor plasma (PPP) + Ca^{2+}	120–240 s (2–4 min)
Citrated platelet-poor plasma + PL + Ca^{2+}	60–85 s
APTT: Citrated platelet-poor plasma + kaolin + PL + Ca^{2+}	21–32 s
PT: Citrated platelet-poor plasma + thromboplastin + Ca^{2+}	11–12 s

Thromboplastin contains FVII and PL

Clotting Factors

The coagulation factors are traditionally identified as roman numerals, based on the order in which they were isolated and identified historically. They circulate in plasma as inactivated and activated (denoted with 'a') proteins. The majority of clotting factors are precursors of proteolytic enzymes known as zymogens that circulate in blood as an inactive form. The activation of each zymogen is depicted by suffixing letter "a" to the Roman numeral identifying that particular zymogen (Fig. 2.1) (Palta et al. 2014). These proenzyme factors undergo a post translational modification (vitamin K dependent gamma (γ) carboxylation of glutamic acid residues) which enables them to bind calcium ions and adopt the suitable conformation required for binding to negatively charged membrane surfaces, such as the activated platelet surface. Deficiency of vitamin K or administration of vitamin K antagonists (*e.g.* warfarin) abolish this carboxylation and it is to date the most powerful anticoagulant drug (Fig. 2.2) (Hoffman and Monroe 2014).

Vitamin K-dependent proteins play a central role in the regulation of blood coagulation. The N-terminal γ-carboxyglutamic acid (Gla) domain anchors the proteins to cell membranes exposing phosphatidylserine in a Ca^{2+} dependent interaction. A



Coagulation Factors	(kDa)	(nM)	(%)
Factor II (Prothrombin)	72	1207 - 1727	40
Factor VII	50	7 - 13	5-10
Factor IX	**56**	**40 - 142**	**10-40**
Factor X	**59**	**101 - 162**	**10-15**
Factor V	**300**	**12 - 42**	**10-15**
Factor VIII	**280**	**0.36 - 1**	**10-40**

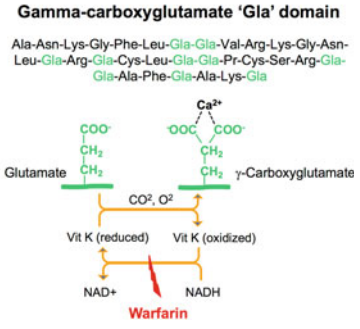


Fig. 2.2 Domains organization and similarities between the Vitamin K dependent coagulation factors in the propagation phase of coagulation. The proenzymes Factors II, VII, IX and X are highly homologous single peptides that comprise of a catalytic/protease domain at the C terminus, followed by a sequence of epidermal growth factor (EGF) or kringle domains, a γ-carboxyglutamic acid (Gla) domain, an activation region (propeptide) and a signal peptide at the N terminus. Upon synthesis, the signal and propeptide domains are cleaved (arrow heads) from the polypeptide chain, exposing the Gla domain to the negatively charged platelet membranes. Further proteolytic activation (arrows) exposes the serine protease site of the enzymes. The table on the right shows the size and concentration of the zymogenes and pro-cofactors in blood (Furie and Furie, 1992)

linear area of Ca^{2+} ions bind internally to the carboxyl groups of the Gla domain inducing a significant structural transition from a largely unfolded to a tightly folded domain capable to bind anionic membranes. The Gla domain-membrane interaction has a critical role in the assembly of the membrane-bound coagulation complexes leading to enhanced catalytic efficiency of the respective proteases. This phenomenon explains the fact that removing Ca^{2+} ions by citrate or EDTA prevents blood from clotting indefinitely (Table 2.1) (Huang et al. 2003; Furie and Furie 1988; Kalafatis and Mann 1997; Kalafatis et al. 1997; Zwaal et al.1998).

Cell Model of Coagulation

Since 2010 a new model of the coagulation process has been proposed—the cell based model that explains some aspects of hemostasis, which a protein-centric (cascade) model does not. The cascade theory is particularly useful for understanding the in vitro coagulation tests, but does not consider the central role of cell-based surfaces in the in vivo coagulation process (Hoffman and Monroe 2001). In the cell-based model, the coagulation process of thrombin generation and fibrin formation is represented on cell surfaces, as a sequence of three distinctive phases (Fig. 2.3):

1. *Initiation phase*: plasma factor (FVII) binds to tissue factor (TF) on subendothelial fibroblasts, auto activates to FVIIa and forms the TF-FVIIa "extrinsic tenase" complex that activates FX to FXa and produces a small amount of thrombin (FIIa).
2. *Propagation and Amplification phase*: The small amount of thrombin (FIIa) generated in the initiation phase activates the platelets. Cleaves FVIII from its carrier-the von Willebrand Factor (vWF). Activates FVIII and FV to FVIIIa and FVa, respectively. Then, it amplifies its own production through the assembly of the FVIIIa-FIXa "intrinsic tenase" complex and the FVa-FXa "prothrombinase" complex on the phosphatidylserine (PS)-rich surfaces of activated platelets. The two complexes work in synergy to amplify and propagate thrombin generation, producing a "thrombin burst".
3. *Fibrin formation*: The thrombin burst is essential for forming cross-linked fibrin (FIIa) from fibrinogen Factor I (FI), by cleaving fibrinogen to soluble fibrin, activating factor XIII (FXII to FXIIIa) is required to cross-link fibrin while concurrently inhibiting fibrinolysis. Calcium (Ca^{2+}) is critical at all stages of thrombin generation and fibrin formation.

It is important to note that from a spatial point of view, everything happens on the injured blood vessel (endothelial) surface. The small amount of thrombin generated by the 'intrinsic' tenase complex also activates the platelets, which includes translocating PS to the platelet surface. These activated platelets together with the vWF, form a platelet plug on the site of injury, which occurs almost simultaneously with the assembly of the tenase and prothrombinase complex.

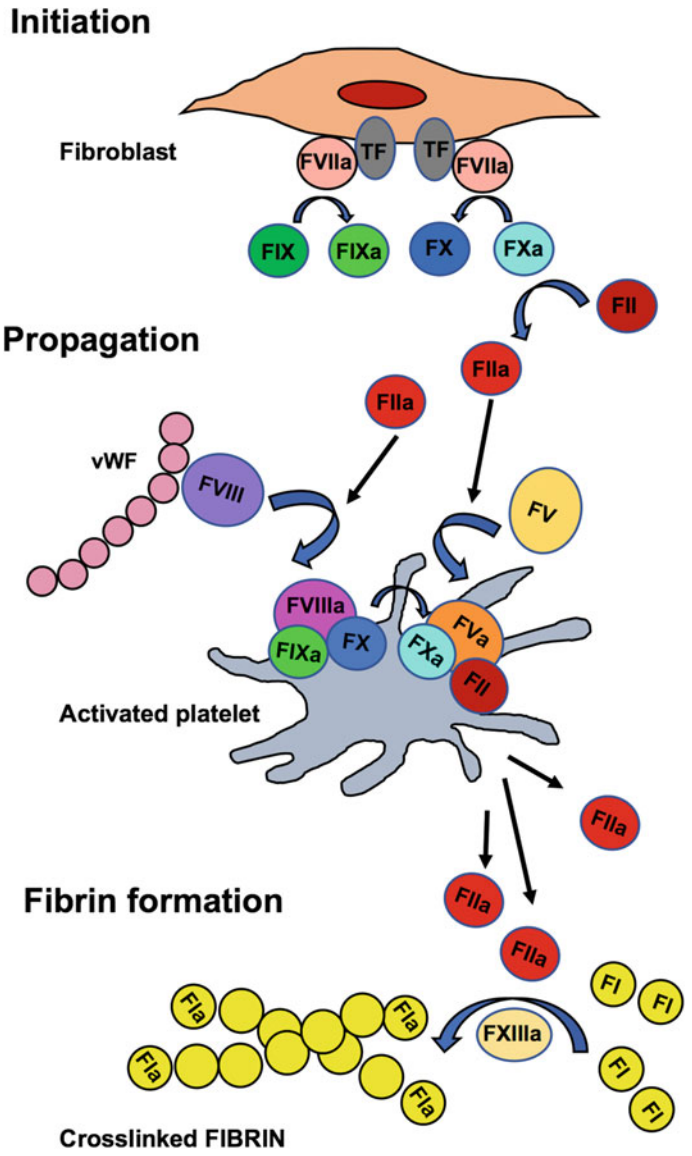


Fig. 2.3 Cell-based model of blood coagulation. Only the proc-coagulant pathways are included in this model, characterized by three phases: Initiation, Propagation and Firbrin formation. The first two phases of the cell model are directly linked to thrombin (FIIa) generation. In the first-initiation phase, -only small (initial) amount of FII are generated on endothelial cells (fibroblast) by the 'intrinsic' tenase complex assmbled on the surface of endothelial cells (fibroblast). In the second-propagation phase, FII generation is increased more than fivefold due to the synergy of the tenase and prothrombinase complex assembled on the activated platelet surface. The thrombin burst generated in the second phase secure accelerated generation of crosslinked fibrin (FIa) in the third-fibrinolysis phase, required to stop fast the blood loss

The models for the coagulation process presented in this introduction include only the pro-coagulant pathways. FVa and FVIIIa are also key players in the anticoagulant pathways, forming membrane-bound complexes with active protein C (APC) and protein S, leading to their inactivation. APC is a vitamin K-dependent serine protease, highly homologous to FVII, FIX an FX. Protein S is a modified vitamin K-dependent serine protease that has a gla-domain facilitating binding to anionic membrane surfaces but does not have a protease domain and proteolytic activity. The FV and FVIII membrane-bound complexes with APC are critical for fine tuning of the coagulation process, more pronounced for FVa, as FVIIIa is intrinsically unstable. In this chapter our focus will be on the pro-coagulant membrane-bound FV and FVIII complexes.

The Tenase and Prothrombinase Complexes

The tenase and prothrombinase complexes are tertiary complexes formed on the PS-rich activated platelet surface in the propagation phase of coagulation. The product from the tenase complex, FXa is the enzyme in the prothrombinase complex that catalyzes the conversion of prothrombin to thrombin. The presence of FVa and FVIIIa within the macromolecular enzyme complexes enhances the respective proteolytic reactions rate by several orders of magnitude, accelerating sufficiently thrombin generation to secure normal hemostasis. The two membrane-bound complexes work in synchronization and are pivotal steps in the coagulation cascade. (Figure 2.4) (Mann et al. 1990).

The serine proteases FIXa and FXa by themselves are capable of activating the respective zymogenes—FX and prothrombin (FII) in solution. The presence of a co-factor and the assembly of the membrane-bound complexes increases the catalytic activity of the enzymes ~ 300,000 times, as illustrated for the prothrombinase complex in Table 2.2.

The significance of the FVIII co-factor function and the tenase complex assembly is further illustrated by the fact that defects or deficiency of FVIII lead to a sever hemorrhagic disorder known as Hemophilia A. Hence FVIII is also known as the anti-hemophilia A factor (AHF) (Cawthern et al. 1998).

Factor V and Factor VIII

Blood coagulation factors V and VIII are two large multidomain glycoproteins that share high structural and functional homology and do not have enzymatic (proteolytic) activity (Mann et al. 1988; Kane and Davie 1988). Factors V and VIII are composed of three structural domains organized in the same order from the N- to the C-terminal: A1-A2-B-A3-C1-C2, with a high degree of amino acid identity (~40%) except for the B domains. The B domain consists of about 900 amino acids and

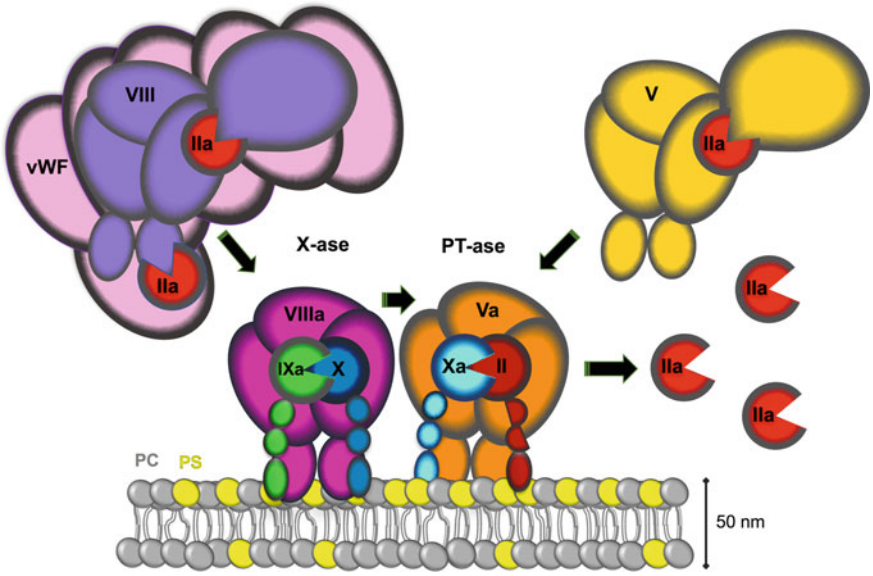


Fig. 2.4 The tenase and prothrombinase complexes. These drawing illustrates the proposed mechanism of the Tenase (X-ase) complex assembly between Factors VIIIa and IXa and the Prothrombinase (PT-ase) complex assembly between Factors Va and Xa on the activated platelet surface rich in PS. Xa is the product of the X-ase and the enzyme of the PT-ase complex. The PT-ase complex activates Prothrombin (II) to Thrombin (IIa). The separate domains of the proteins are sown as ovals/sphere. The membrane and protein complexes are drawn to scale

Table 2.2 Clotting times for FII (thrombin) activation by FXa. PL indicates phospholipids

Clotting factors	Rate of FII generation	Clotting times
II, Xa	1	3,024,000 s (35 days)
II, Xa, PL, Ca^{2+}	350	8640 s (2.4 h)
II, Xa, PL, Va, Ca^{2+}	21,000	150 s (2.5 min)
I, Xa, Platelets, Va, Ca^{2+}	302,400	10 s

The components of the prothrombinase complex In vitro are: FII, FXa, FVa, PL and Ca2 + . In vivo platelets serve the role of PL

is rich in *N*-and *O*-glycosylation sites, 25 for the Factor V and 18 for the FVIII, heavily glycosylated with N-linked oligosaccharides and are highly flexible (Fig. 2.5) (Nesheim et al. 1979; Rosing et al. 1980). The significance of the B domain uniqueness has not been fully understood. The A domains are about 330 amino acid and have 30% identity between each other and the three A domains of Curuloplasmin, a Cu-binding protein that circulates freely in blood. The C domains are about 150 amino acids and are highly homologous to the Discoidin 1 protein (Pemberton et al. 1997; Saenko et al. 2002).

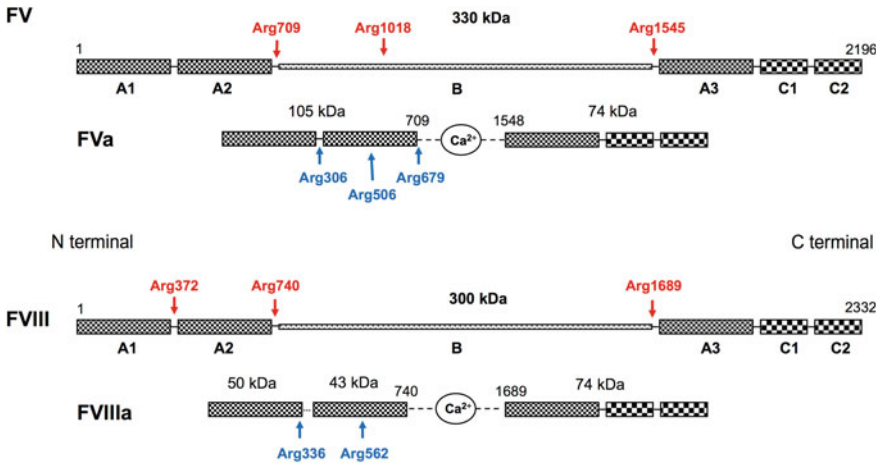


Fig. 2.5 Domains organization of Factor V and Factor VIII. FV and FVIII are expressed as single chain six domains proteins of ~ 300 kDa. The three A domains are homologous (~ 35% sequence identity) and contain the FX and FXa binding sites, respectively. The A2 and A3 domains are separated by a unique B domain that is not required for function. The C domains are also homologous (~ 40% sequence identity) and contain the membrane-binding sites. Upon proteolytic activation by Thrombin (red arrows) the B domain is excised and the FVa and FVIIIa active forms present as non-covalently bound heterodimers of a heavy chain (A1-A2 domains) and a light chain (A3-C1-C2) domains stabilize electrostatically by Ca^{2+} ions. The A1 and A2 domains in FVIII, further separate by proteolysis at Arg372, leading to the final active form presenting a heterotrimer of the A1 and A2 domains holding together solely by hydrophobic interactions and the light chain stabilized by both hydrophobic and electrostatic interactions. The blue arrows present the APC cleavage sites, leading to inactivation of FV and FVIII

The activated forms, FVIIIa and FVa act as co-factors to the serine proteases FIXa and FXa respectively, when assembled in a membrane-bound complex with a zymogen substrate FX and prothrombin, respectively (Fig. 2.4). These complexes are functional only when formed in the presence of calcium ions and anionic phospholipid surfaces, such as the PS rich activated platelet membrane. Calcium is needed to stabilize the Gla-domains of the vitamin K-dependent serine proteases so they can efficiently bind to the negatively charge activated platelet membrane (Huang et al. 2004). Calcium is also needed for the single calcium binding sites in FVIII and FV that are required for stabilizing the heavy and light chain of the proteins after activation (Fig. 2.5) but have not shown to be required for membrane-binding (Stoilova-McPhie et al. 2013). The assembly of the FVIIIa/FIXa and FVa/FXa complexes increases the Vmax of the FX and prothrombin activation ~5 orders of magnitude (Mann et al. 1988; Kane and Davie 1988; Nesheim et al. 1984a).

Although the knowledge of factor VIII and factor V and their regulation has progressed significantly over the last two decades, the molecular mechanisms by which their co-factor function leads to accelerated substrate cleavage remains to be fully defined. The difficulties in having homogeneous preparations of these factors have been overcome by the advent of recombinant DNA technologies and protein

purification procedures. The dynamics and complex interactions of these factors with their multiple partners (ions, membranes, activators and inhibitors) renders them a change for structural and functional studies to date. There has not yet been a consensus on the molecular models of FVIIIa and FVa within the membrane-bound tenase and prothrombinase complexes, despite existing crystal structures for FVIII and FV, the interacting proteases—FIXa and FXa and zymogen substrates—FX and prothrombin, respectively. Understanding the mechanism by which the structure of these factors morphs to fulfill their function is the endeavor of modern hematology (Bos and Camire 2010).

Factor V

Factor V (FV) is a large (Mr ~ 330 kDa), multidomain (A1-A2-B-A3-C1-C2), single chain glycoprotein that circulates in blood at a concentration of ~ 20 nM (~ 10 μg/ml). At physiological plasma concentrations, the pro-cofactor FV cannot assemble or function in the prothrombinase complex. Proteolytic processing of the B domain is an absolute requirement for the expression of cofactor function. Thrombin is the main physiological activator of FV and cleaves three peptide bonds (Arg709, Arg1018, and Arg1545) within the B domain that leads to the B domain removal and FV activation (Fig. 2.5). The active from—FVa, presents a heterodimer composed of a heavy chain (A1-A2; Mw = 105 kDa) and a light chain (A3-C1-C2; Mw = 71–74 kDa), which are associated through electrostatic forces supported by Ca^{2+} ions (Fig. 2.5) (Jenny et al. 1987; Mann and Kalafatis 2003; Suzuki et al. 1982; Esmon et al. 1973; Esmon 1979; Nesheim et al. 1984b; Stoilova-McPhie et al. 2008).

Deficiency of FV is associated with a rare bleeding disease (1:1,000,000) parahemophilia. Most mutations of FV are associated with venous thrombosis due to resistance to cleavage by APC, leading to the FVa remaining active and accelerated thrombin generation through the prothrombinase complex. The most common of these mutations is known as the FV Leiden (Asselta et al. 2006; Sun et al. 1994; Bos and Camire 2010).

Factor VIII

Factor VIII (FVIII) is expressed as a large (Mr ~ 330 kDa), heavily glycosylated multidomain (A1-A2-B-A3-C1-C2) proteins that circulates in vivo, as a complex with the von Willebrand factor (vWF)—a large (Mr ~ 2MDa) multidomain protein (Figs. 2.3 and 2.4). Once activated by thrombin it exists as a heterodimer resulting from limited proteolysis at the B-A3 junction and at additional sites in the B domain. This heterodimer consists of a variably sized heavy chain (A1-A2-B; 200-290 kDa) and a light chain (A3-C1-C2; ~ 80 kDa) that are non-covalently associated and correspond to the FVIII form purified from blood (plasma derived, pd) or expressed

in vitro (recombinant FVIII) (Fig. 2.5). The heavy chain interaction with the light chain serves various important roles in FVIII physiology, as it has been reported to stabilize the heterodimeric structure of FVIII, as well as to prevent proteolysis by FXa and APC. The FVIII heterodimer is an inactive pro-cofactor and it is subjected to further proteolysis to gain full cofactor activity (Fay 2004; Fay 2006). Factor VIII can be proteolytically activated by both thrombin and FXa. Thrombin cleaves three peptide bonds at Arg372, Arg740, and Arg1689, thereby generating FVIIIa, which is a heterotrimer composed of the A1 (Mr ~ 50 kDa) and A2 (Mr ~ 43 kDa) domains held together by hydrophobic forces, and the light chain (A3-C1-C2; Mr ~ 73 kDa) (Fig. 2.5). Once activated, FVIIIa biological activity decays over a short period of time (~ 6 min in vitro) due to the A2 domain dissociation from A1/A3-C1-C2 domains, a mechanism that contributes to the regulation of FVIIIa cofactor activity (Vehar et al. 1984; Fay et al. 1991; Fay and Smudzin 1992; Lollar and Parker 1991; Fay 2004; Fay 2006; Lenting et al. 1998).

Defect or deficiency of FVIII result in Hemophilia A, an X chromosome genetic disorder affecting 1:5000 males (Graw et al. 2005). Recombinant FVIII was launched as drug for Hemophilia A in 1992 and it was one of the first recombinant proteins approved for treatments in humans. Multiple commercial recombinant FVIII forms exist to date, that facilitate significant structural and functional studies (Antonarakis 1995; Oldenburg et al. 2004).

Despite the similarities in the structure and function of FV and FVIII, the molecular mechanisms of activation have been found to differ substantially. The inactive pro-cofactor is stable when purified from blood or expressed. The FV does not have a transient heterodimer form before the final activation step to FVa. Activated FV (FVa) forms a stable heterodimer that can be characterized structurally and functionally (Fig. 2.5 (Stoylova et al.1994; Stoilova-McPhie et al. 2008). The FVa inactivation is achieved by forming a membrane-bound complex with APC/ProteinS and FVIII. The inactivated FVai lacks its A2 domain that contains the main interactions sites with the serine protease FXa. The structure of the FVai (Adams et al. 2004) and consequent structures of FVa (not publicly available, because of ownership by pharmaceutical companies) are highly homologous between themselves and to the resolved FVIII structures (RMSD between the Cα atoms is ~ 1.5 Å). The mechanism of activation of FVIII is quite different. The stable form is a heterodimer of variable size which is an intermediate product between the stable complex of the pro-cofactor, FVIII with the vWF and the stable complex between the activated FVIII (FVIIIa) and the FIXa within the membrane-bound tenase complex (Fay 2006). To further stabilize the inactive FVIII heterodimer, B domain deleted (BDD-FVIII) constructs were generated producing FVIII homodimers that are more stable and homogenous in solution and can be studies structurally and functionally. The deletion of the B-domain increases the yield but does not affect the FVIII in vitro and in vivo activity (Parameswaran et al. 2005; Kessler et al. 2005). Cleavage between the A1 and A2 domains at position Arg372 is critical for the pro-coagulant activity of FVIII. The precise mechanism by which cleavage at Arg372 facilitates the transition to the active cofactor state remains to be determined (Nicolaes and Dahlback 2002; Bos and Camire 2010).

The Factor VIII Structures

The macromolecular organization of the FVIII domains in their membrane-bound form and their synchronization are the driving force behind FVIII activity and the assembly of the FVIIIa-FIXa complex on the activated platelet surface. The presence of multiple membrane-binding sites at the FVIII light chain interface is due to the flexibility of the molecule and specifically the flexible link between the C1 and C2 domains lacking a stable hydrophobic macromolecular interface (Stoilova-McPhie et al. 2013). While several structures of FVIII constructs organized in 3D crystals were resolved at ~ 3.6 to 4Å resolution by X-ray crystallography (Ngo et al. 2008; Shen et al. 2008; Svensson et al. 2008; Leksa et al. 2017; Smith et al. 2020), a different membrane-bound domain organization has been proposed from our cryo-EM studies (Fig. 2.6) (Stoilova-McPhie et al. 2002; Stoilova-McPhie et al. 2013; Miller et al. 2014; Dalm et al. 2015).

The structural basis of FVIII activation and the FVIIIa-FIXa complex assembly strongly depend on the FVIII domain organization and geometry. Until we have direct structural evidence for the FVIIIa-FIXa complex assembly on a membrane mimicking the activated platelets surface, the functional implications of FVIII structure will not be completely understood. The macromolecular forces that drive the

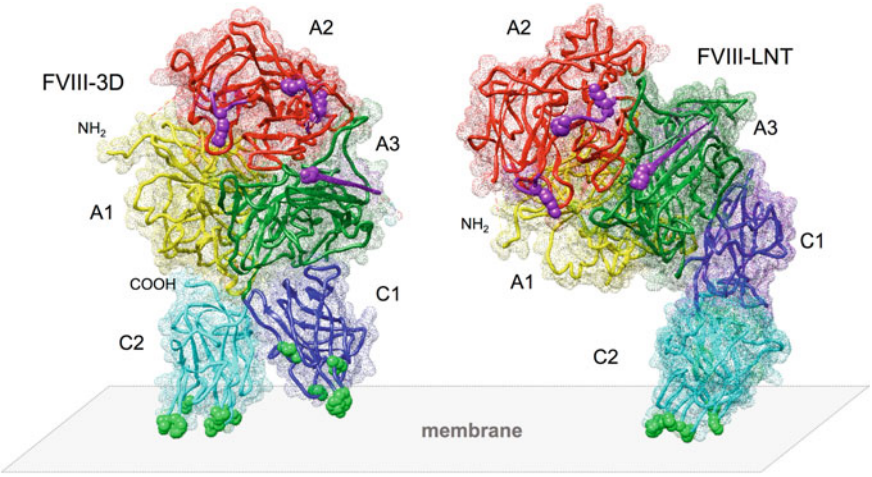


Fig. 2.6 FVIII structure organized in 3D (FVIII-3D) and helical crystals (FVIII-LNT). FVIII-3D is the BDD-FVIII structure as organized in 3D crystals (PDB: 3CDZ) resolved at ~ 4 Å resolution by X-ray crystallography (Ngo et al. 2008). The five FVIII domains are shown as ribbons and colored as follows: A1-yellow, A2-red, A3-green, C1-blue, C2-cyan. The FVIII membrane-binding residues identified on the C1 and C2 domains are shown in green. The FVIII residues identified at the FVIII-FIXa interface on the A2 and A3 domains are shown in purple. The N terminal (Ala1) is indicated with NH_2. FVIII-LNT is the FVIII structure based on the helical organization of the FVIII light chain bound to lipid nanotubes (3J2S) resolved by cryo-EM at 15 Å resolution (Stoilova-McPhie et al. 2013; Dalm et al. 2015). The lipid membrane surface is denoted with a grey rectangle

FVIII activation and the FVIIIa-FIXa complex assembly on a PS-rich membrane surface are fundamental for the successful design of membrane-mimetics modulating blood hemostasis at the FVIII level.

FVIII Structure as Helically Organized on Lipid Nanotubes

Self-organized lipid nanotubes (LNT) can be formed from Galactosylceramide (GC) lipids under specific conditions. These nanotubes are functionalized with specific ligands incorporated into the lipid bilayer. Any protein that can bind to these ligands can be subjected to helical organization and further structure determination by cryo-EM (Wilson-Kubalek et al. 2010; Egelman 2010). Single bilayer GC-LNT rich in PS, have been optimized in our group for the helical organization of FVIII. The geometry and lipid composition of these LNTs mimics the tubular protrusions (pseudopodia) formed by blood platelets upon their activation (Parmenter and Stoilova-McPhie 2008; Parmenter et al. 2008; Wassermann et al. 2010). Three recombinant FVIII proteins, two human and one porcine, were helically organized and the structure resolved by cryo-EM at 15Å and 14.5Å, respectively. All helical reconstructions showed that the membrane-bound FVIII macromolecules formed tightly organized and stable membrane-bound dimers (Fig. 2.7) (Stoilova-McPhie et al. 2013; Stoilova-McPhie et al. 2014; Miller et al. 2014; Dalm et al. 2015; Stoilova-McPhie 2017). The FVIII membrane-bound dimer organization was further confirmed by electron tomography of the helically organized FVIII on LNT (Miller et al. 2014; Dalm et al. 2015) and for FVIII bound to nanodiscs with the same lipid composition as the LNT (Stoilova-McPhie et al. 2014; Grushin et al. 2015; Grushin and Stoilova-McPhie 2017). The FVIII dimer interface was defined by rigid body docking of two FVIII-LNT FVIII structures (Fig. 2.6) within the cryo-EM map of the segmented dimer as shown on Fig. 2.7C. Our studies show that the main interaction interface between the two FVIII molecules is at the FVIII-heavy chain and predominantly at the level of the A2 domains, as only 8 of the 33 residues identified at the FVIII dimer interface within a cutoff of 8.7Å belong to the A1 domain (Fig. 2.8). Fourteen of these residues differ between the human and porcine sequence and eight are known Hemophilia A mutations (Stoilova-McPhie et al. 2014; Stoilova-McPhie 2017). The dimeric FVIII membrane-bound model proposed from our cryo-EM studies is supported by data reported for dimeric membrane-bound assembly of other coagulation factors, such as human FVIII, FIXa, bovine FVa and FXa (Stoylova, Mann and Brisson 1994; Stoilova-McPhie et al. 2014; Chattopadhyay et al. 2009; Miller et al. 2014).

The Supertenase Complex

The FVIIIa-FIXa complex assembly is regulated through high-affinity protein-lipid and protein-protein interactions. The dissociation constants (Kd) for the FVIIIa and FIXa membrane interactions are ~ 1 nM and for the FVIIIa-FIXa interaction ~ 15 nM

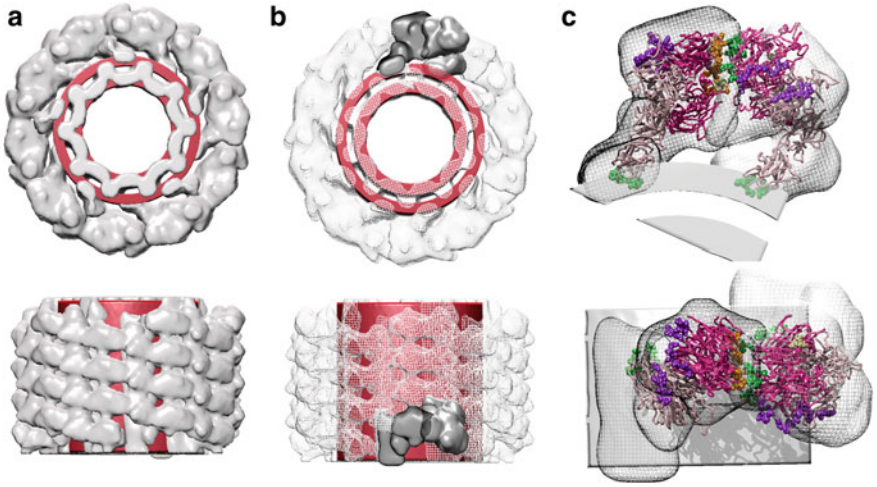


Fig. 2.7 Helical reconstruction of pFVIII-LNT. a. Helical reconstruction of porcine FVIII-LNT (light grey) from 10,433 segments at 14.5Å resolution (0.5 FSC) superimposed with the 3D reconstruction of a lipid nanotube (LNT) from 2000 segments (red). The helical parameters: twist – $\Delta\varphi = 35.5^{o}$ and rise - $\Delta z = 36.0$ Å converged after 200 IHRSR cycles. **b**. Segmentation showing the asymmetric unit cell (dark grey). **c**. Rigid body fitting of the FVIII-LNT structure within the EM map of the FVIII dimer. The FVIII-Light chain is shown in light pink and the heavy chain in hot pink ribbons. The FVIII-FVIII dimer interface is shown with green and orange spheres respectively. The FVIII-membrane interface is shown with green spheres and the FVIII-FIXa interface with magenta spheres. The top row shows views along the helical axis and the bottom row, views perpendicular to the helical axis (LNT membrane). The inner diameter of the LNT is 2.0 nm

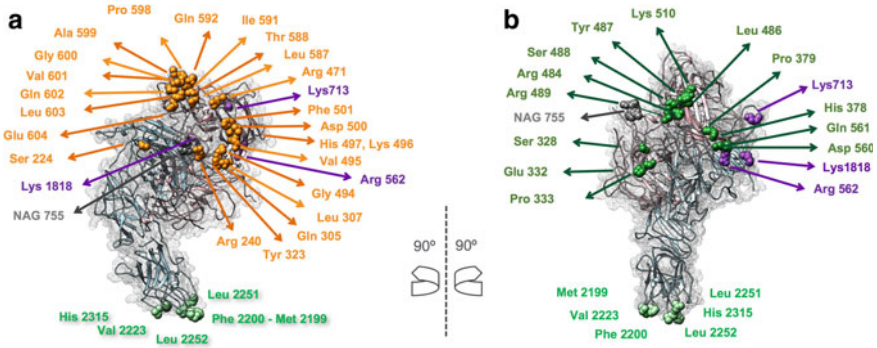


Fig. 2.8 The FVIII dimer interface. The FVIII dimer interface is shown parallel to the membrane surface. **a**. Interface of the monomer on the right in Fig. 2.7c facing the FVIII dimer interface. **b**. Interface facing the monomer on the left in Fig. 2.7c facing the FVIII dimer interface. There are 33 interactions between 22 amino acid residues from the 'left' monomer interface, shown with orange spheres with 13 residues from the 'right' monomer interface, shown as green spheres within a radius less than 7.8 Å. 14 of these residues differ between the human and porcine FVIII sequence, and 8 residues have known Hemophilia A (HA) mutations

(Ahmad, London and Walsh 2003). The extended FVIIIa-FIXa interface consists of four distinctive interaction sites, which have been identified through a number of functional studies assessing the inhibition by peptides, antibodies, direct mutagenesis, specific point mutations and cross-linking studies: (i) FVIIIa-A2 domain (558-565)—FIX-protease domain (helix 330-338) (ii) FVIIIa-A2 domain (707-713)—FIXa-protease domain (helix 301-303), (iii) FVIIIa-A3 domain (1811-1818) with FIXa-EGF1 domain and (iv) FVIIIa-C2 domain (2228-2240) with FIXa-Gla domain (Fig. 2.9) (Fay et al. 1994; Bajaj et al. 2001; Griffiths et al. 2013; Kolkman et al. 1999; Lenting et al. 1996a; Lenting et al. 1996b; Mathur and Bajaj 1999; Blostein et al. 2003; Soeda et al. 2009). Based on this knowledge combined with the resolved FVIII-2D (PDB: 3J2Q), FVIII-3D (3CDZ) and the FIXa (1PFX) structures, two very different models of the Tenase complex have been proposed (Brandstetter et al. 1995; Ngo et al. 2008; Autin et al. 2005; Stoilova-McPhie et al. 2002). In the first model FVIII (3CDZ) binds to FIXa in solution then undergoes a significant domain reorganization, such that all of the FVIII-LC domains interact with the membrane (Ngo et al. 2008) In the second model, FVIII (3J2Q) first undergoes significant domain reorganization upon binding to the membrane in a conformation suitable to bind FIXa

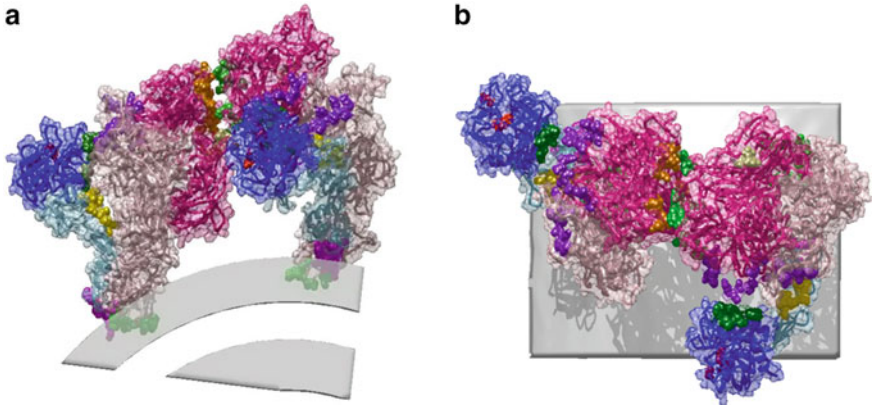


Fig. 2.9 The FVIIIa-FIXa complex. A, B are views along and perpendicular to the LNT surface. The FVIII-light chain is colored in light pink and the heavy chain in hot pink. The residues at the FVIII monomer-monomer interface are shown as green and orange spheres, respectively. The C2 domain membrane-binding residues are shown as light green spheres. The FVIII residues on the A2 and A3 domains identified to bind to the FIXa are shown was purple spheres. The FIXa (1PFX) heavy chain (protease domain) is colored in dark blue and the light chain holding the two EGF like domains in light blue. The FIXa membrane-binding Gla domain is shown as magenta spheres. The Asp39, His57 and Cys195 residues forming the FIXa serine protease active site are shown as red spheres. The identified primary FIXa binding sites for FVIIIa on the Protease domain are shown with green spheres: Lys132-Phe133-Gly134-Ser135 and Pro161-Leu162-Val163-Asp164-Arg165-Ala166-Thr167-Cys168-Leu169-Arg170. The secondary FIXa-FVIIIa interaction sites on the FIXa-FGF domains are shown with yellow spheres: Lys80-Asn81–Glu83-Leu84-Asp85-Ala86-Thr87-Cys88-Asp89-Ile90. The residues from the membrane-binding Gla domain are shown as cyan spheres. The LNT's bilayer structure as defined by cryo-EM is shown as a light grey surface (Stoilova-McPhie, 2017).

in the FVIIIa-FIXa complex (Stoilova-McPhie et al. 2002). Resolving the FVIII and FVIIIa membrane-bound structures as assembled on LNT by cryo-EM will be decisive in building a functional model of the FVIIIa-FIXa complex and understanding the critical role played by FVIII in hemostasis.

The dimeric FVIIIa-FIXa (supertenase) complex was modelled with the membrane-bound porcine FVIII dimer as a template, the porcine FIXa crystal structure (1PFX) and the identified FVIIIa-FIXa interaction sites imposing similar constraints as the ones used for the previous Tenase models. First, the FVIIIa and FIXa molecules were oriented such as to face each other at the appropriate height above the membrane (Mutucumarana et al. 1992). Second, the two molecules were rotated and translated in the membrane plane such as to accommodate the four identifed FVIIIa-FIXa interactions sites and secure that the membrane-binding FIXa-Gla and FVIIIa-C2 domains interact correctly with the lipid bilayer. Third, the FIXa active site (Asp39, His57, Ser195 catalytic triad) needs to be exposed to the substrate—FX.

Further, flexible fitting will be required to refine the model based on subnanometer resolution cryo-EM structures. None of the existing models, however, can accommodate all four FVIIIa-FXa interaction sites. The proposed dimeric model has the advantage that the substrate, FX can be positioned correctly on the membrane surface, such as it is stabilized by interaction with the FVIII-FX binding site identified on the A1 domain and accessible to the FIXa proteolytic/active site. The dimeric/multimeric organization of the tenase and prothrombinase complexes will also justify a structural synchronization of the protein function needed for more than five order of magnitude increase in the proteolytic activity for the respective enzymes (FIXa and FXa).

Structural Basis of FVIII Activation-Future Perspectives

Despite the considerable efforts have been made over the last two decades to define the mechanism of FVIII and FV activation, the structural basis of their co-factor functions is poorly understood. (Mann and Kalafatis 2003; Fay 2004). The active form of FV, FVa, is stable and we have previously obtained 2D crystals from membrane-bound bovine FVa (Stoylova et al. 1994). The active form of FVIII, FVIIIa is much more volatile and unstable for direct structural studies. The porcine FVIII (pFVIII) expressed and activated in our lab had sufficient stability and quantity for functional and structural studies by cryo-EM. The purity and integrity of the pFVIII and pFVIIIa were assessed as described by Lollar et al. (1993). The FVIII and FVIIIa were both helically organized on LNT at the same lipid composition and lipid to protein ratio, as previously described (Miller et al. 2014; Dalm et al. 2015). The current 3D reconstruction of pFVIII-LNT and pFVIIIa-LNT reached 12Å resolution at Fourier Shell Correlation coefficient, FSC = 0.5 (Fig. 2.10). Further segmentation of the helical volumes confirmed the dimeric organization of both porcine FVIII and porcine FVIIIa when bound to the LNT surface. This first membrane-bound FVIIIa structural data confirmed our previous observation from membrane-bound 2D

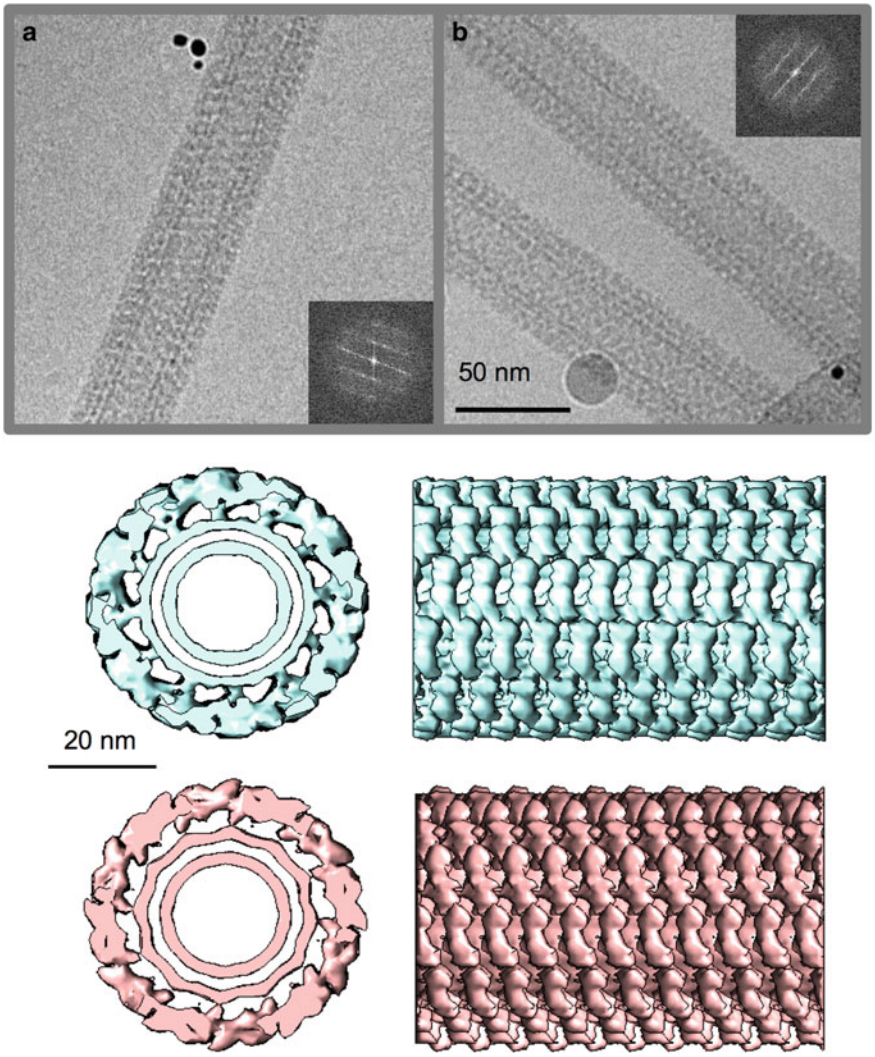


Fig. 2.10 Cryo-EM micrographs of pFVIII (A) and pFVIIIa (B) helically organized on LNT (1024 × 1024 pixels at 1.4 Å/pix resolution). The insets are Fourier transform of the helical tubes (512 × 512 pixels) showing the layer lines corresponding to the helical organization. Helical reconstruction of pFVIII-LNT (in blue) and pFVIII-LNT (in pink) calculated from 951 and 944 helical segments (512 × 512 pixels) respectively, as described in (Dalm et al. 2015)

crystals of plasma derived FVIIIa (unpublished), that there are no significant differences in the oligomeric organization of the proteins after activation. This finding confirmed our hypothesis, that significant domain reorganization of FVIII occurs

upon membrane-binding, not activation. Activation of the pro-cofactor is subtle structurally, optimizing solely the FVIIIa-FIXa interface following proteolysis of FVIII at Arg372.

It is important to note that all crystallographic structural studies of serine protease so far, including FIXa and FXa have been carried out with active site-inhibited constructs lacking the membrane-bounding Gla domain. Our FVIII cyo-EM studies have been focused on the porcine BDD-FVIII construct that has been proven to be more stable than the human analogue. At the same time porcine FVIII has been successfully used for therapies for Hemophilia A, as it has been shown to form active complexes with human FIXa (Lollar et al. 1993). Our working hypothesis for the assembly of the tenase complex is that first the dimeric FVIII-FIXa complex is assembled that requires significant reorganization of the FVIII domain upon membrane-binding followed by dimerization (Dalm et al. 2015). The FVIII activation happens for the membrane-bound FVIII form, restructuring the FVIIIa-FIXa interface to increase singificantly FIXa proteolytic activity (Fig. 2.11).

Conclusions

The prothrombinase complex is the last step of the propagation phase of coagulation impacting thrombin generation and the clotting process as shown in Fig. 2.4 and Table 2.2. For that reason, most pharmaceutical interventions in the last 20 years have been directed towards modulators of the FXa active site. Hence more than 100 crystal structure of FXa with different inhibitors exist in the protein structure data bases (RSCB.org). A finer and refined modulation of the pro-coagulant pathways can be achieved by targeting the co-factors and specifically the FVIIIa/FIXa and the FVIII-FVIII dimer interface within the membrane-bound tenase complex. These studies are however only in an initial stage. Several crystal structures of FVa exist that are not publicly available except for the FVai (Adams et al. 2004), and a structure of a soluble, not membrane-bound prothrombinase complex, has been resolved by X-ray crystallography (Lechtenberg et al. 2013; Csencsits-Smith, Grushin and Stoilova-McPhie 2015). The exact mechanism of the FVa-FXa interaction is not yet fully characterized that makes it difficult to target with novel therapeutics. The situation is more complex with the FIXa and tenase complex, as the FVIIIa form is highly unstable and has not so far been structurally resolved with any of the existing direct structural approaches: X-ray crystallography, cryo-EM and NMR. In our lab, we conducted cryo-EM studies of FVIII bound to nanodiscs and tested our constructs in hemophilic A mice followed by a proposal for nanodrugs based on our design for treatment of Hemophilia A. Present achievement in both the nanotechnology and Cryo-EM fields will soon lead to our full understanding of the FVIII and FV membrane-bound complexes assembly and regulation.

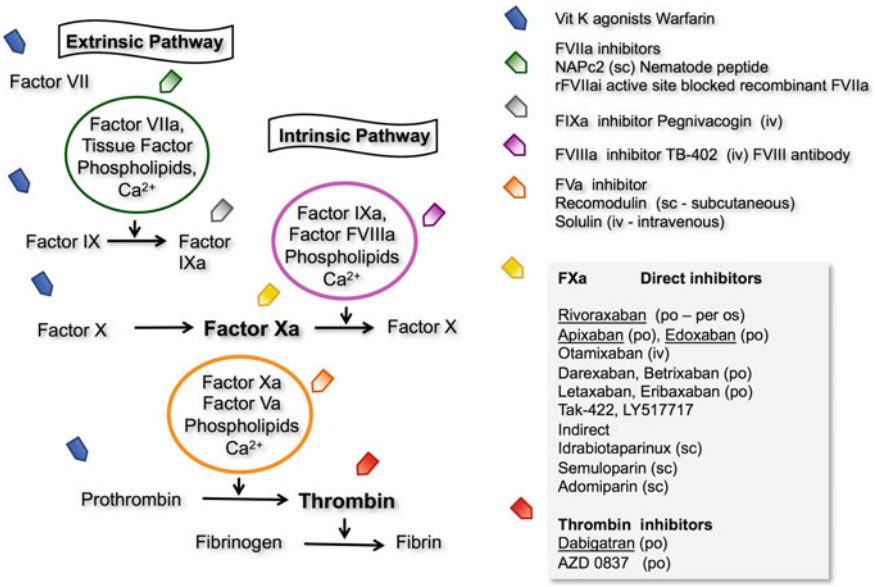


Fig. 2.11 Oral anticoagulant (non-Vitamin K Antagonist, NOAC) targeting the three membranefigound complexes in the pro-coagulant pathways of the clotting process. So far mainly the serine proteases and specifically FXa have been the target for modulation of the coagulation process (Fig. 2.11). These new pharmaceutics, known as oral anticoagulant (po) or non-vitamin K antagonists (NOAC) can inhibit a single protease, are sufficiently small to reach the active site and their binding is reversible. The dosage and side effects of these oral coagulants however, can be severe. Resolving the structural basis and mechanism of FVIII and FV co-factor functions within the membrane-bound tenase and prothrombinase complexes will open avenues for pharmaceuticals interventions at a new level of specificity and efficacy, providing efficient cures for coagulation disorders (Hoffman and Monroe, 2017). In this chapter, we have focused on the FVIII and the tenase (FVIII/FIXa) membrane bound complex assembly and macromolecular organization. The protocols developed for these studies can be extended to the prothrombinase (FVa/FXa) complex and the FVa and FVIIIa anti-coagulant membrane-bound complexes with APC. Resolving the functional structure of the components of these complexes in a membrane environment and at atomic resolution by cryo-EM will help elucidate their function. This will benefit not only drug discovery, but will advance our scientific understanding of structure-function relarionship at the membrane interface. It will open novel avenues for membrane-based nanotechnology designs that can modulate protein activity though the assembly of intricate macromolecular complexes

Acknowledgements SSM acknowledges all her co-authors, former group members, mentors, and supervisors. SSM acknowledges the Sealy Center for Structural Biology at the University of Texas Medical Branch at Galveston, TX, USA, and the National Center for Molecular Imaging at Baylor College of Medicine, Houston, TX, and the respective directors: Professors B. Montgomery Pettitt and Wah Chiu (now in Stanford) for continuous support. SSM acknowledges the British Heart Foundation for project grant PG/04/070, the American Heart Association fort eh national scientist development grant 10SDG3500034, The University of Texas Medical Branch for startup funds and the John Sealy Foundation at Galveston, TX for their bridging grant.

References

Adams TE, Hockin MF, Mann KG, Everse SJ (2004) The crystal structure of activated protein C-inactivated bovine factor Va: implications for cofactor function. Proc Natl Acad Sci U S A 101(24):8918–8923

Ahmad SS, London FS, Walsh PN (2003) The assembly of the factor X-activating complex on activated human platelets. J Thromb Haemost 1(1):48–59

Antonarakis SE (1995) Molecular genetics of coagulation factor VIII gene and hemophilia A. Thromb Haemost 74(1):322–328

Asselta R, Tenchini ML, Duga S (2006) Inherited defects of coagulation factor V: the hemorrhagic side. J Thromb Haemost 4(1):26–34

Autin L, Miteva MA, Lee WH, Mertens K, Radtke KP, Villoutreix BO (2005) Molecular models of the procoagulant factor VIIIa-factor IXa complex. J Thromb Haemost 3(9):2044–2056

Bajaj SP, Schmidt AE, Mathur A, Padmanabhan K, Zhong D, Mastri M, Fay PJ (2001) Factor IXa:factor VIIIa interaction. helix 330-338 of factor ixa interacts with residues 558–565 and spatially adjacent regions of the a2 subunit of factor VIIIa. J Biol Chem 276(19):16302–16309

Blostein MD, Furie BC, Rajotte I, Furie B (2003) The Gla domain of factor IXa binds to factor VIIIa in the tenase complex. J Biol Chem 278(33):31297–31302

Bos MH, Camire RM (2010) Blood coagulation factors V and VIII: Molecular Mechanisms of Procofactor Activation. J Coagul Disord 2(2):19–27

Brandstetter H, Bauer M, Huber R, Lollar P, Bode W (1995) X-ray structure of clotting factor IXa: active site and module structure related to Xase activity and hemophilia B. Proc Natl Acad Sci USA 92(21):9796–9800

Cawthern KM, van 't Veer C, Lock JB, DiLorenzo ME, Branda RF, Mann KG (1998) Blood coagulation in hemophilia A and hemophilia C. Blood 91(12):4581–92

Chattopadhyay R, Iacob R, Sen S, Majumder R, Tomer KB, Lentz BR (2009) Functional and structural characterization of factor Xa dimer in solution. Biophys J 96(3):974–986

Csencsits-Smith K, Grushin K, Stoilova-McPhie S (2015) Binding of factor VIII to lipid nanodiscs increases its clotting function in a mouse model of hemophilia A. J Blood Disord Transfus 6(6):325

Dalm D, Galaz-Montoya JG, Miller JL, Grushin K, Villalobos A, Koyfman AY, Schmid MF, Stoilova-McPhie S (2015) Dimeric organization of blood coagulation factor viii bound to lipid nanotubes. Sci Rep 5:11212

Davie EW, Ratnoff OD (1964) Waterfall sequence for intrinsic blood clotting. Science 145(3638):1310–1312

Egelman EH (2010) Reconstruction of helical filaments and tubes. Methods Enzymol 482:167–183

Esmon CT (1979) The subunit structure of thrombin-activated factor V. Isolation of activated factor V, separation of subunits, and reconstitution of biological activity. J Biol Chem 254(3):964–973

Esmon CT, Owen WG, Duiguid DL, Jackson CM (1973) The action of thrombin on blood clotting factor V: conversion of factor V to a prothrombin-binding protein. Biochim Biophys Acta 310(1):289–294

Fay PJ (2004) Activation of factor VIII and mechanisms of cofactor action. Blood Rev 18(1):1–15

Fay PJ (2006) Factor VIII structure and function. Int J Hematol 83(2):103–108

Fay PJ, Smudzin TM (1992) Characterization of the interaction between the A2 subunit and A1/A3-C1-C2 dimer in human factor VIIIa. J Biol Chem 267(19):13246–13250

Fay PJ, Haidaris PJ, Smudzin TM (1991) Human factor VIIIa subunit structure. Reconstruction of factor VIIIa from the isolated A1/A3-C1-C2 dimer and A2 subunit. J Biol Chem 266(14):8957–8962

Fay PJ, Beattie T, Huggins CF, Regan LM (1994) Factor VIIIa A2 subunit residues 558-565 represent a factor IXa interactive site. J Biol Chem 269(32):20522–20527

Furie B, Furie BC (1988) The molecular basis of blood coagulation. Cell 53(4):505–518

Furie B, Furie BC (1992) Molecular and cellular biology of blood coagulation. N Engl J Med 326(12):800–806

Gale AJ (2011) Continuing education course #2: current understanding of hemostasis. Toxicol Pathol 39(1):273–280
Graw J, Brackmann HH, Oldenburg J, Schneppenheim R, Spannagl M, Schwaab R (2005) Haemophilia A: from mutation analysis to new therapies. Nat Rev Genet 6(6):488–501
Griffiths AE, Rydkin I, Fay PJ (2013) Factor VIIIa A2 subunit shows a high affinity interaction with factor IXa: contribution of A2 subunit residues 707–714 to the interaction with factor IXa. J Biol Chem 288(21):15057–15064
Grushin KWM, Stoilova-McPhie S (2017) Reversible stacking of lipid nanodiscs for structural studies of clotting factors. Nanotechnology Reviews
Grushin K, Miller J, Dalm D, Stoilova-McPhie S (2015) Factor VIII organisation on nanodiscs with different lipid composition. Thromb Haemost 113(4):741–749
Hoffman M, Monroe DM 3rd (2001) A cell-based model of hemostasis. Thromb Haemost 85(6):958–965
Hoffman M, Monroe DM (2014) Reversing targeted oral anticoagulants. Hematology Am Soc Hematol Educ Program 1:518–523
Hoffman M, Monroe DM (2017) Impact of non-vitamin K antagonist oral anticoagulants from a basic science perspective. Arterioscler Thromb Vasc Biol 37(10):1812–1818
Huang M, Rigby AC, Morelli X, Grant MA, Huang G, Furie B, Seaton B, Furie BC (2003) Structural basis of membrane binding by Gla domains of vitamin K-dependent proteins. Nat Struct Biol 10(9):751–756
Huang M, Furie BC, Furie B (2004) Crystal structure of the calcium-stabilized human factor IX Gla domain bound to a conformation-specific anti-factor IX antibody. J Biol Chem 279(14):14338–14346
Jenny RJ, Pittman DD, Toole JJ, Kriz RW, Aldape RA, Hewick RM, Kaufman RJ, Mann KG (1987) Complete cDNA and derived amino acid sequence of human factor V. Proc Natl Acad Sci USA 84(14):4846–4850
Kalafatis M, Mann KG (1997) Factor VLeiden and thrombophilia. Arterioscler Thromb Vasc Biol 17(4):620–627
Kalafatis M, Egan JO, van 't Veer C, Cawthern KM, Mann KG (1997) The regulation of clotting factors. Crit Rev Eukaryot Gene Expr 7(3):241–80
Kane WH, Davie EW (1988) Blood coagulation factors V and VIII: structural and functional similarities and their relationship to hemorrhagic and thrombotic disorders. Blood 71(3):539–555
Kessler CM, Gill JC, White GC 2nd, Shapiro A, Arkin S, Roth DA, Meng X, Lusher JM (2005) B-domain deleted recombinant factor VIII preparations are bioequivalent to a monoclonal antibody purified plasma-derived factor VIII concentrate: a randomized, three-way crossover study. Haemophilia 11(2):84–91
Kolkman JA, Lenting PJ, Mertens K (1999) Regions 301-303 and 333-339 in the catalytic domain of blood coagulation factor IX are factor VIII-interactive sites involved in stimulation of enzyme activity. Biochem J 339(Pt 2):217–221
Langdell RD, Wagner RH, Brinkhous KM (1953) Effect of antihemophilic factor on one-stage clotting tests; a presumptive test for hemophilia and a simple one-stage antihemophilic factor assy procedure. J Lab Clin Med 41(4):637–647
Lechtenberg BC, Murray-Rust TA, Johnson DJ, Adams TE, Krishnaswamy S, Camire RM, Huntington JA (2013) Crystal structure of the prothrombinase complex from the venom of Pseudonaja textilis. Blood 122(16):2777–2783
Leksa NC, Chiu PL, Bou-Assaf GM, Quan C, Liu Z, Goodman AB, Chambers MG, Tsutakawa SE, Hammel M, Peters RT, Walz T, Kulman JD (2017) The structural basis for the functional comparability of factor VIII and the long-acting variant recombinant factor VIII Fc fusion protein. J Thromb Haemost 15(6):1167–1179
Lenting PJ, van Mourik JA, Mertens K (1998) The life cycle of coagulation factor VIII in view of its structure and function. Blood 92(11):3983–3996

Lenting PJ, Christophe OD, Maat H, Rees DJ, Mertens K (1996a) Ca2 + binding to the first epidermal growth factor-like domain of human blood coagulation factor IX promotes enzyme activity and factor VIII light chain binding. J Biol Chem 271(41):25332–25337

Lenting PJ, van de Loo JW, Donath MJ, van Mourik JA, Mertens K (1996b) The sequence Glu 1811-Lys1818 of human blood coagulation factor VIII comprises a binding site for activated factor IX. J Biol Chem 271(4):1935–1940

Lollar P, Parker ET (1991) Structural basis for the decreased procoagulant activity of human factor VIII compared to the porcine homolog. J Biol Chem 266(19):12481–12486

Lollar P, Fay PJ, Fass DN (1993) Factor VIII and factor VIIIa. Methods Enzymol 222:128–143

Macfarlane RG (1964) An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. Nature 202:498–499

Mann KG, Kalafatis M (2003) Factor V: a combination of Dr Jekyll and Mr Hyde. Blood 101(1):20–30

Mann KG, Jenny RJ, Krishnaswamy S (1988) Cofactor proteins in the assembly and expression of blood clotting enzyme complexes. Annu Rev Biochem 57:915–956

Mann KG, Nesheim ME, Church WR, Haley P, Krishnaswamy S (1990) Surface-dependent reactions of the vitamin K-dependent enzyme complexes. Blood 76(1):1–16

Mathur A, Bajaj SP (1999) Protease and EGF1 domains of factor IXa play distinct roles in binding to factor VIIIa. Importance of helix 330 (helix 162 in chymotrypsin) of protease domain of factor IXa in its interaction with factor VIIIa. J Biol Chem 274(26):18477–18486

Miller J, Dalm D, Koyfman AY, Grushin K, Stoilova-McPhie S (2014) Helical organization of blood coagulation factor VIII on lipid nanotubes. J Vis Exp (88)

Mutucumarana VP, Duffy EJ, Lollar P, Johnson AE (1992) The active site of factor IXa is located far above the membrane surface and its conformation is altered upon association with factor VIIIa. A fluorescence study. J Biol Chem 267(24):17012–17021

Nesheim ME, Taswell JB, Mann KG (1979) The contribution of bovine Factor V and Factor Va to the activity of prothrombinase. J Biol Chem 254(21):10952–10962

Nesheim ME, Foster WB, Hewick R, Mann KG (1984a) Characterization of Factor V activation intermediates. J Biol Chem 259(5):3187–3196

Nesheim ME, Tracy RP, Mann KG (1984b) "Clotspeed," a mathematical simulation of the functional properties of prothrombinase. J Biol Chem 259(3):1447–1453

Ngo JC, Huang M, Roth DA, Furie BC, Furie B (2008) Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. Structure 16(4):597–606

Nicolaes GA, Dahlback B (2002) Factor V and thrombotic disease: description of a janus-faced protein. Arterioscler Thromb Vasc Biol 22(4):530–538

Oldenburg J, Ananyeva NM, Saenko EL (2004) Molecular basis of haemophilia A. Haemophilia 10(Suppl 4):133–139

Palta S, Saroa R, Palta A (2014) Overview of the coagulation system. Indian J Anaesth 58(5):515–523

Parameswaran R, Shapiro AD, Gill JC, Kessler CM, Investigators HR (2005) Dose effect and efficacy of rFVIIa in the treatment of haemophilia patients with inhibitors: analysis from the hemophilia and thrombosis research society registry. Haemophilia 11(2):100–106

Parmenter CD, Stoilova-McPhie S (2008) Binding of recombinant human coagulation factor VIII to lipid nanotubes. FEBS Lett 582(12):1657–1660

Parmenter CD, Cane MC, Zhang R, Stoilova-McPhie S (2008) Cryo-electron microscopy of coagulation Factor VIII bound to lipid nanotubes. Biochem Biophys Res Commun 366(2):288–293

Pemberton S, Lindley P, Zaitsev V, Card G, Tuddenham EG, Kemball-Cook G (1997) A molecular model for the triplicated A domains of human factor VIII based on the crystal structure of human ceruloplasmin. Blood 89(7):2413–2421

Quick AJ (1950) Estimation of prothrombin time. Can Med Assoc J 63(1):89

Rosing J, Tans G, Govers-Riemslag JW, Zwaal RF, Hemker HC (1980) The role of phospholipids and factor Va in the prothrombinase complex. J Biol Chem 255(1):274–283

Saenko EL, Ananyeva NM, Tuddenham EG, Kemball-Cook G (2002) Factor VIII—novel insights into form and function. Br J Haematol 119(2):323–331

Shen BW, Spiegel PC, Chang CH, Huh JW, Lee JS, Kim J, Kim YH, Stoddard BL (2008) The tertiary structure and domain organization of coagulation factor VIII. Blood 111(3):1240–1247

Smith IW, d'Aquino AE, Coyle CW, Fedanov A, Parker ET, Denning G, Spencer HT, Lollar P, Doering CB, Spiegel PC Jr (2020) The 3.2 A structure of a bioengineered variant of blood coagulation factor VIII indicates two conformations of the C2 domain. J Thromb Haemost 18(1):57–69

Soeda T, Nogami K, Nishiya K, Takeyama M, Ogiwara K, Sakata Y, Yoshioka A, Shima M (2009) The factor VIIIa C2 domain (residues 2228–2240) interacts with the factor IXa Gla domain in the factor Xase complex. J Biol Chem 284(6):3379–3388

Stoilova-McPhie S (2017) Lipid nanotechnologies for structural studies of membrane-associated clotting proteins by cryo-electron microscopy. Nanotechnol Rev

Stoilova-McPhie S, Lynch GC, Ludtke S, Pettitt BM (2013) Domain organization of membrane-bound factor VIII. Biopolymers 99(7):448–459

Stoilova-McPhie S, Grushin K, Dalm D, Miller J (2014) Lipid nanotechnologies for structural studies of membrane-associated proteins. Proteins 82(11):2902–2909

Stoilova-McPhie S, Villoutreix BO, Mertens K, Kemball-Cook G, Holzenburg A (2002) 3-Dimensional structure of membrane-bound coagulation factor VIII: modeling of the factor VIII heterodimer within a 3-dimensional density map derived by electron crystallography. Blood 99(4):1215–1223

Stoilova-McPhie S, Parmenter CD, Segers K, Villoutreix BO, Nicolaes GA (2008) Defining the structure of membrane-bound human blood coagulation factor Va. J Thromb Haemost 6(1):76–82

Stoylova S, Mann KG, Brisson A (1994) Structure of membrane-bound human factor Va. FEBS Lett 351(3):330–334

Sun X, Evatt B, Griffin JH (1994) Blood coagulation factor Va abnormality associated with resistance to activated protein C in venous thrombophilia. Blood 83(11):3120–3125

Suzuki K, Dahlback B, Stenflo J (1982) Thrombin-catalyzed activation of human coagulation factor V. J Biol Chem 257(11):6556–6564

Svensson LA, Bondensgaard K, Norskov-Lauritsen L, Christensen L, Becker P, Andersen MD, Maltesen MJ, Rand KD, Breinholt J (2008) Crystal structure of a prolactin receptor antagonist bound to the extracellular domain of the prolactin receptor. J Biol Chem 283(27):19085–19094

Vehar GA, Keyt B, Eaton D, Rodriguez H, O'Brien DP, Rotblat F, Oppermann H, Keck R, Wood WI, Harkins RN, Tuddenham EG, Lawn RM, Capon DJ (1984) Structure of human factor VIII. Nature 312(5992):337–342

Wassermann GE, Olivera-Severo D, Uberti AF, Carlini CR (2010) Helicobacter pylori urease activates blood platelets through a lipoxygenase-mediated pathway. J Cell Mol Med 14(7):2025–2034

Wilson-Kubalek EM, Chappie JS, Arthur CP (2010) Helical crystallization of soluble and membrane binding proteins. Methods Enzymol 481:45–62

Zwaal RF, Comfurius P, Bevers EM (1998) Lipid-protein interactions in blood coagulation. Biochim Biophys Acta 1376(3):433–453

Chapter 3
An Overview of Dps: Dual Acting Nanovehicles in Prokaryotes with DNA Binding and Ferroxidation Properties

Sunanda Margrett Williams and Dipankar Chatterji

Abstract DNA binding proteins under starvation (Dps) are proteins belonging to the ferritin family with the capacity for DNA binding, in addition to iron storage and ferroxidation. Present only in the prokaryotes, these multifaceted proteins have been assigned with a number of roles, from pathogenesis to nucleoid condensation and protection. They have a significant role in protecting the cells from free radical assaults, indirectly by sequestration of iron and by directly binding to the DNA. Due to their symmetry, stability and biomineralization capacity, these proteins have ever increasing potential applications in biotechnology and drug delivery. This chapter tries to bring together all these aspects of Dps in the view of current understanding and older perspectives by studies of our group as well as other experts in the field.

Keywords DNA binding proteins · Ferritin family · Iron oxidation · Atomic force microscopy · Nanovehicles

Introduction

Dps (DNA-binding protein from starved cells) are homo-dodecamers belonging to the ferritin family which can bind DNA, carry out ferroxidation and store iron. These proteins are present only in members of eubacteria and archaea. In nature, iron starvation and oxidative stress are two hurdles that bacteria must circumvent to establish growth. To overcome the paradox of iron and oxidative stress, as free iron contributes to further oxidative stress by its participation in the Fenton reaction (Fenton 1894), iron is stored in a non-toxic form by Dps proteins.

S. M. Williams (✉)
Institute of Structural and Molecular Biology, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom
e-mail: s.williams@mail.cryst.bbk.ac.uk

D. Chatterji (✉)
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India
e-mail: dipankar@iisc.ac.in

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_3

Dps also binds DNA in a sequence independent manner and brings about its condensation. Thus, it is classed as a nucleoid associated protein (NAP) similar to HU, IHF, H-NS etc. Dps seems to play a multi-faceted role in bacterial cells, as it performs a number of diverse roles in various organisms. This section attempts to give a general introduction from the discovery of Dps to a description of its role in protecting cells under various stresses and its rapid degradation in the cells during logarithmic growth phase through the N-end rule degradation pathway.

Discovery of DNA Binding Proteins Under Starvation (Dps)

The Kolter group (Almiron et al. 1992) while studying gene expression of starved *Escherichia coli* cells discovered the over expression of a certain DNA binding protein whose expression depended on a stationary phase specific transcription factor σS. In negative staining electron microscopy, it had a double layered hexameric ring like appearance and formed two dimensional arrays with DNA (Fig. 3.1a, b). Later a homolog discovered in nitrogen starved *Synechococcus* sp. PCC7942 named DpsA showed the presence of a carboxyl domain homologous to the C termini of bacterioferritins (Pena and Bullerjahn 1995), giving first indications of its iron binding property. Subsequently, a dodecameric ferritin-like protein isolated from the gram-positive *Listeria innocua* able to sequester around 500 iron atoms and oxidise iron, was shown to be a Dps-like protein (Bozzi et al. 1997).

More comprehensive information about the Dps structure which gave an experimental basis for its functions, was obtained after the crystal structure of Dps from *E. coli* was solved by Hogle, Kolter and colleagues. The X-ray crystal structure showed

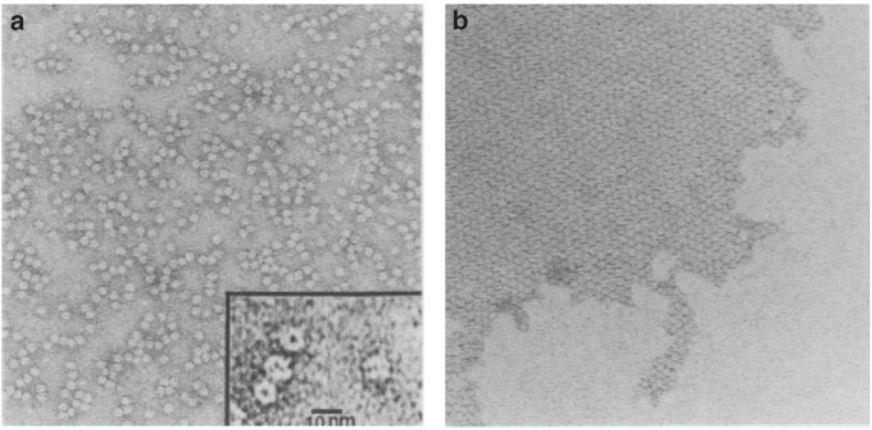


Fig. 3.1 *Negative stain electron microscopic analysis of* **a** Dps protein, inset represents a higher magnification showing ring-like structures and **b** Dps-DNA two dimensional crystalline arrays. Images reproduced with permission from Genes & Dev. 1992, 6: 2646–2654

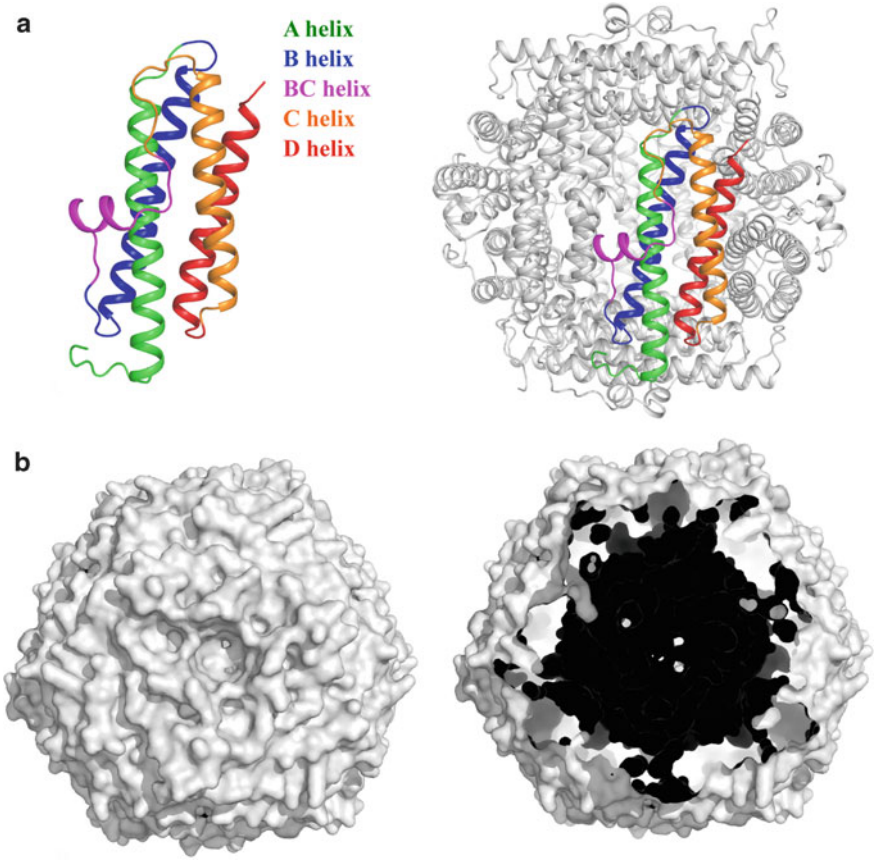


Fig. 3.2 *X-ray structure of E. coli Dps (PDB ID: 1dps)* **a** Alpha helical monomer showing A, B, C, D and BC helices; the right panel shows assembly as a dodecamer **b** Surface representation of a dodecamer with the outer view (left) and a section showing the hollow interior (right)

Dps monomers to have a ferritin-like fold assembling into 12-mers with tetrahedral (23) symmetry, enclosing a hollow core with pores at the three-folds (Grant et al. 1998) (Fig. 3.2a, b). The structure of another Dps homolog from *L. innocua* gave the first direct evidence of a ferroxidase center indicating a structural basis of ferroxidase activity (Ilari et al. 2000). X-ray structures of more Dps homologs revealed these features to be consistent within the family.

Iron Oxidation and Storage

Iron, one of the most abundant metals on earth, plays a vital role in the biology of living organisms. It is a cofactor in many redox reactions and can be found in

the active sites of a number of enzymes. Iron can act as an electron source and a sink, and has a central role in various cellular processes like metabolism, respiration, nucleotide synthesis, gene regulation etc. Iron starvation or a lack of this valuable metal in the environment poses an adverse threat to existence.

Free iron can exist in two states, ferrous (Fe^{2+}) and ferric (Fe^{3+}) in organisms. The ferrous state is soluble and readily assimilated in cells, but the ferric state is highly insoluble and unavailable for physiological functions. But the more bioavailable ferrous form has a tendency to oxidise to ferric thereby limiting its bioavailability and also generating hydroxyl radicals in this process. H_2O_2, a product of oxidative respiration reacts with ferrous iron to generate hydroxyl free radicals which are toxic to cells, by a process called the Fenton reaction (Eq. 3.1) (Martinez and Kolter 1997).

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^{\bullet} + OH^{-} \tag{3.1}$$

Thus, it is essential that iron is stored in a readily accessible but non-toxic form in cells. Ferritin, a family of iron storage proteins are the primary iron storage compartments in vertebrates, invertebrates, plants and microorganisms. They are structurally adept at storing iron in their hollow protein shells in a state that is available and non-deleterious. The ferritin family of proteins includes ferritins and other ferritin-like proteins such as the haem containing bacterioferritins and Dps which can also bind DNA (Andrews 2010).

Ferritins and other ferritin-like proteins have a hollow central cavity where iron is oxidised and stored as ferric oxide. Based on their capacity to store iron, the ferritin-like proteins can be also be widely categorised as the 24-meric maxiferritins which include ferritins and bacterioferritins and 12-meric miniferritins which are the Dps. The miniferritins or Dps have been recognised only in eubacteria and archaea, unlike the maxiferritins which are present in bacteria, archaea and the eukaryotes. The maxiferritins due to their bigger protein shells which consequently enclose a larger cavity, can store >4000 iron atoms per protein. Miniferritins on the other hand can store only around 500 iron atoms per dodecamer. Hydrogen peroxide is the preferred reagent for iron oxidation in miniferritins, unlike maxi ferritins where dioxygen is the major electron acceptor (Zhao et al. 2002).

In Dps, iron incorporation is a multi-step process of Fe (II) binding, Fe(II) oxidation, nucleation and growth of the mineral core, in a reaction pathway similar to classical ferritins (Zhao et al. 2002).

Fe(II) Binding (Phase 1)

24 Fe(II) binds at the 12 di-iron binding ferroxidation sites in the protein

$$2Fe^{2+}\,pz \rightarrow \left[Fe(II)_2 - P\right]_{FS}^{Z+2} + 2H^{+} \tag{3.2}$$

where $[Fe(II)_2–P]^{Z+2}{}_{FS}$ represents a di-Fe(II)-protein complex at each of the 12 putative ferroxidase sites. For more information about the ferroxidase centre and di-iron binding sites see section *Conserved ferroxidase active site,* below.

Fe(II) Oxidation (Phase 2)

Two di-nuclear ferroxidase sites bring about the rapid pairwise oxidisation of two Fe (II) by one molecule of H_2O_2. Thus, two Fe (II) are oxidised per H_2O_2 reduced, preventing the generation of hydroxyl radicals through Fenton chemistry (Eq. 3.1).

$$\left[\mathrm{FE(II)_2 - P}\right]^{Z+2}_{FS} + H_2O_2 + H_2O \rightarrow \left[\mathrm{Fe(III)_2O_2(OH) - P}\right]^{Z-1}_{FS} + 3H^+ \quad (3.3)$$

Nucleation and Growth of Mineral Core (Phase 3)

A ferric core of approximately 500 Fe(III) is formed inside the Dps protein shell, according to the mineralisation equation with a 2 Fe(II)/H_2O_2 stoichiometry:

$$2Fe^{2+} + H_2O_2 + 2H_2O \rightarrow 2Fe(III)OOH(core) + 4H^+ \quad (3.4)$$

A visual schematic of the ferroxidation pathway is represented in Fig. 3.14. In Dps, ferroxidation can also take place with O_2 as an oxidant, but the reaction happens at a much slower rate.

DNA Protection by Direct and Indirect Modes

DNA is often called the blueprint of life. It is in the interest of cell survival to devise highly efficient DNA protection mechanisms to offset the lethal effects caused by DNA lesions. But during conditions of nutrient depletion the cell cannot afford to repair DNA through energy extravagant DNA defence pathways. In starved *E. coli* cells for example, Dps accumulates to very large amounts and forms the major component of chromatin in late stationary phase bacteria (Martinez and Kolter 1997). Dps binds DNA and compacts it during stress, thereby exposing fewer unprotected regions to reactive agents. Oxidation of Fe(II) to Fe(III) prevents the ferrous form from participating in Fenton reaction (Eq. 3.1) quenching further production of free radicals deleterious to DNA. The ferroxidation activity also results in the reduction of toxic peroxides by utilization of H_2O_2 as an oxidant. Dps also has a role in the regulation of the expression of certain proteins under conditions of stress.

A bimodal protection of DNA by Dps was seen in homolog from *Mycobacterium smegmatis*, MsDps1 (Gupta and Chatterji 2003). This homolog has two oligomeric forms, dimer and dodecamer. The dimeric form has properties of ferroxidation only, whereas the dodecamer could store iron, carry out ferroxidation and bind DNA. The dimer could still protect DNA against the assault of free radicals, as it has intact ferroxidase sites and thus can carry out ferroxidation. In the earlier publication (Gupta and Chatterji 2003) the lower oligomer was identified as a trimer, whereas recently through mass spectrometry it has been correctly assigned as a dimer (Williams et al. 2017).

The DNA binding property and protection can be assessed by simple biochemical assays, which monitor DNA retardation or degradation in the presence and absence of Dps proteins. DNA retardation of pUC19 plasmid by Dps1 from *M. smegmatis* can be visualised on an agarose gel (Fig. 3.3a). Protection of DNA can be observed by treating DNA with H_2O_2, where the DNA with bound Dps is shielded (Fig. 3.3b). DNA protection from DNaseI digestion is also accorded by Dps (Fig. 3.3c), which prevents access to DNA by these degrading enzymes (Gupta and Chatterji 2003).

Thus, Dps renders protection by directly binding to DNA and also indirectly by preventing the generation of toxic free radicals through its ferroxidation property. A more detailed description of the mechanisms involved in DNA binding and protection are discussed below.

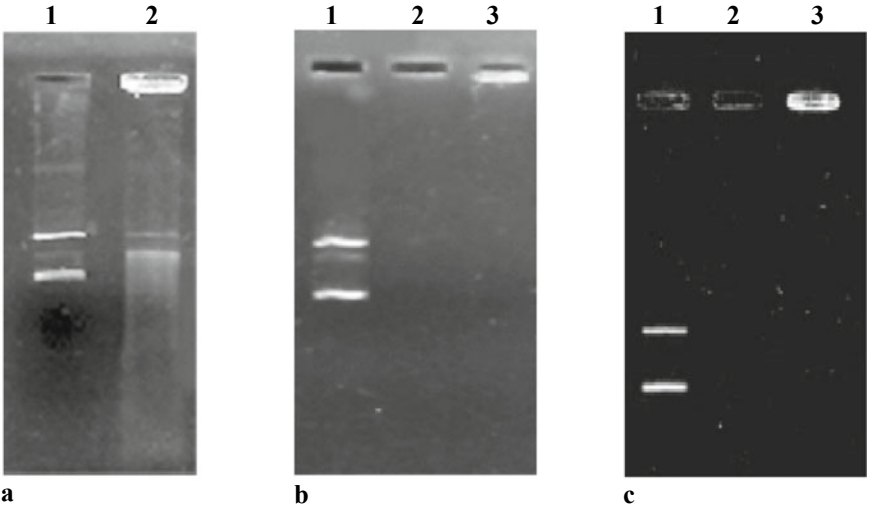


Fig. 3.3 *Gel based assays for DNA protection by M.smegmatis Dps* **a** *Gel retardation of DNA by Dps.* Free pUC19 DNA (lane 1) and Dps-DNA complex shows retardation of DNA in the well (lane 2). **b** *DNA protection from peroxide damage.* pUC19 DNA (lane 1) pUC19 DNA treated initially with 50 μM $FeSO_4$ followed by 5 mM H_2O_2 for 5 min degrades the DNA completely (lane 2) and Dps-DNA complex treated with 50 μM $FeSO_4$ followed by 5 mM H_2O_2 for 5 min where the DNA is protected (lane3). **c** *DNA protection from DNaseI digestion.* pUC19 DNA (lane 1) DNaseI treated pUC19 DNA (lane 2) Dps-DNA complex protected from DNaseI digestion (lane3). Images reproduced from J. Biol. Chem. 2003, 278: 5235–5241

Dps Across the Prokaryota

The role of Dps in bacterial cells can differ widely, while retaining some of its classic properties, like DNA binding and iron storage. *E. coli* Dps is one of the best studied Dps homolog and exhibits most of the typical properties of Dps family proteins. *E. coli* Dps is additionally known to protect the cells from UV and gamma radiations, iron/copper toxicity, thermal stress, acid and base shock in addition to oxidative stress (Nair and Finkel 2004). To better appreciate the numerous homologs of Dps in bacteria and their varied roles, we have summarised some of the better studied Dps proteins in prokaryotes.

Dps homologs like the *L. innocua* Dps do not bind DNA but can protect it against the deleterious combination of Fe^{2+} and H_2O_2 as evidenced by DNA cleavage assays (Su et al. 2005). The plant pathogen, *Agrobacterium tumifaciens* has a Dps with a truncated N-terminus which is fixed on the protein surface and not available for DNA binding, unlike the long and flexible N-terminus of *E. coli* Dps which has DNA binding property. In *A. tumifaciens*, Dps acts in concert with catalase A to counteract toxic peroxides that are major components of plant defense systems against invading bacteria (Ceci et al. 2003).

The *Streptococcus mutans* Dps homolog named Dpr (Dps-like peroxide resistance) confers peroxide resistance, helping the bacteria to survive in aerobic conditions. This organism is a principal causative agent of dental caries and lacks a respiratory chain and the enzyme catalase responsible for H_2O_2 elimination (Yamamoto et al. 2002). Similarly, *S. pyogenes* does not produce catalase, but can grow in aerobic conditions due to the presence of Dpr. Both these streptococcal Dpr proteins do not bind DNA, but provide protection from H_2O_2 stress through nullifying the Fenton reaction and sequestration and oxidation of iron (Tsou et al. 2008). The swine pathogen *S. suis* Dpr also protects DNA from hydrogen peroxide stress by removing free Fe^{2+} from the cytosol, without any apparent DNA binding ability (Kauko et al. 2006). *Bacillus anthracis* BaDps1 and BaDps2 acts as a pair to carry out iron sequestration and H_2O_2 depletion respectively, without binding to DNA (Schwartz et al. 2010). *B. cereus* has three Dps homologs Dps1, Dps2 and Dps3 of which only the latter shows DNA binding property (Shu et al. 2013).

The Dps-1 from *Deinococcus radiodurans* an organism best known for its unusual resistance to ionising radiations, fails to protect DNA from hydroxyl radical mediated cleavage, due to continuous release of stored iron from the protein, although it can bind DNA. This release of iron from the protein core was attributed to the presence of a distinct iron-exit channel. The other Dps homolog in the same organism, Dps-2, has a signal peptide which directs it to the perimeter of *D. radiodurans* cells. Dps-2 shows weak affinity to DNA but displays robust iron storage and ferroxidation. These distinct functional properties and localisation suggests that Dps-1 may participate as a NAP, whereas Dps-2 protects against exogenous reactive oxygen species (ROS) due to its location in the perimeter (Reon et al. 2012).

Lactococcus lactis DpsA and DpsB form Dps:DNA complexes which appear as non-crystalline aggregates in electron micrographs, instead of the ordered crystalline

arrays seen in Dps: DNA complexes of *E. coli* and other homologs. The disordered N-terminus in *E. coli* and *M. smegmatis* is required for formation of crystalline array with Dps. The N-terminal region of *L. lactis* on the other hand has an ordered helical conformation. Also, *L. lactis* Dps was unable to sequester iron due to a preponderance of positively charged residues near the iron entry pores, which occludes iron due to charge repulsion (Stillman et al. 2005). The non-pathogenic mycobacterial species *M. smegmatis* has two Dps homologs, MsDps1 and MsDps2. These are differently regulated in the cells, where MsDps1 is expressed in the stationary phase and is controlled by the extra cellular sigma factors sigF and sigH, whereas MsDps2 is constitutively expressed by RNA polymerase under the control of sigma factors sigA and sigB (Chowdhury et al. 2007). *Streptomyces coelicolor*, a soil dwelling gram positive bacterium, has three Dps homologs, DpsA, DpsB and DpsC out of which DpsA is over-expressed during osmotics shock and DpsC is a tailless protein that does not oligomerise (Facey et al. 2009).

The Dps from *Helicobacter pylori* is a neutrophil activating protein HP-NAP, and drives inflammation in allergic bronchial asthma in humans and mice (Kottakis et al. 2008). HP-NAP of *H. pylori* does not bind DNA, but addition of Fe^{2+} was shown to promote DNA binding. *Borrelia burgdorferi* Dps NapA, is also a neutrophil activating protein and elicits host immune response (Li et al. 2007). *Microbacterium arborescens* Dps termed AAH (acyl amino acid hydrolase) displays additional catalytic activities such as amide hydrolysis and synthesis (Pesek et al. 2011). *Porphyromonas gingivalis* Dps, PgDps binds heme through its conserved cysteine residue and confers resistance to heme toxicity (Gao et al. 2012). *Campylobacter jejuni* Dps has a role in biofilm formation and cecal colonization in poultry, in addition to protecting DNA in the acidic conditions of the host digestive tract and inside phagolysosomes (Theoret et al. 2012).

The role of Dps in extremophiles like *D. radiodurans*, which survive high ionising radiation has been mentioned above. In thermophiles like *Streptococcus thermophilus*, Dps functions as a cold shock protein (Nicodeme et al. 2004). An archaeal Dps protein DpsA from the halophilic marine archaea *Halobacterium salinarum*, renders protection for aerobic survival (Zeth et al. 2004). The Dps from *Sulfolobus solfataricus* a thermophilic archaeon, has an active site similar to manganese catalases, which can bind Mn^{2+} in addition to Fe^{2+} (Hayden and Hendrich 2010).

In pathogenic bacteria, Dps helps the cells to evade host defense pathways. In *Candidatus Legionella jeonii*, DpsX protects the cells from oxidative stress generated by the phagocytic activities of the host (Park et al. 2006). In *Legionella pneumophilia*, DpsL is upregulated during iron starvation and helps the pathogen to circumvent iron limiting conditions (Yu et al. 2009). *Bacteriodes fragilis*, an opportunistic pathogen of the GI tract causes infection when the organism escapes the anaerobic colon to aerobic sites like the peritoneum with high O_2 concentrations. Here, resistance to oxidative stress is important and is involved in the initiation and persistence of infection. One of the proteins involved in detoxification of oxygen radicals is Dps and signifies its role in oxidative stress response of aerobic and facultative organisms (Rocha et al. 2003). Dps in *Helicobacter hepaticus*, a bacterium associated with chronic hepatitis

and hepatocellular carcinoma in mice, plays an important role in protecting DNA from oxidative damage (Hong et al. 2006). In *Salmonella enterica*, Dps helps to resist iron dependent killing by hydrogen peroxide and promotes virulence and survival of the organism in macrophages (Pacello et al. 2008). MrgA, a Dps homolog in *Staphylococcus aureus*, also resists phagocytic killing by macrophages (Gaupp et al. 2012).

The cyanobacterium *Nostoc punctiforme* features five Dps proteins, NpDps 1–5, each having distinct physiological differences and cell-specific expression (Howe et al. 2018). In photosynthetic cyanobacteria like *Synechococcus sp* strain 7942, Dps localizes preferentially to the cytosolic side of thylakoids to provide iron to iron-rich photosystem I (Durham and Bullerjahn 2002). Anabena sp. PCC 7120 has two Dps homologs, which plays a role in alleviating oxidative stress due to environmental changes like temperature and light in addition to its usual function of iron storage and oxidation (Narayan et al. 2010). *Thermosynechococcus elongatus* Dps homologs, Dps-Te and DpsA-Te, inhibits Fenton mediated damage and this is of special significance in the protection of DNA and photosystems I and II from hydrogen peroxide mediated oxidative damage, as it lacks the catalase gene (Franceschini et al. 2006). In the N_2 fixing cyanobacterium *Trichodesmium erythraeum*, Dps_{tery} protects the cells from oxidative damage (Castruita et al. 2006).

N-end Rule Degradation

We have seen that Dps is an important multifunctional protein in bacteria which accumulates to a variety of stresses in the cells, such as nutrient and oxidative stress. Dps is typically over expressed during the stationary phase of the cells and rapidly degraded during logarithmic phase by ClpX/ClpP (Stephani et al. 2003) and ClpS/ClpA/ClpP proteases (Schmidt et al. 2009). The N-terminus of Dps has ClpX and ClpS recognition motifs which promotes degradation by ClpP and ClpA/ClpP proteases, respectively. In *E. coli* Dps, ClpX recognition motif is the N terminus without the first methionine residue *i.e.* Dps2-167, is degraded by ClpXP proteolysis. ClpS targets the N-terminally truncated Dps variant with the first five residues missing, Dps6-167 harbouring the destabilizing Leu6 at the N-terminus.

The N-terminus of Dps in *E.coli* not only serves as a protease recognition motif, but is also important for DNA binding (Ceci et al. 2004). It is proposed that the N-terminal tail of Dps is sequestered during stationary phase when it binds to DNA and this could protect it from degradation. Generation of Dps2-167 or Dps6-167 through degradation of unprotected N- termini could initiate the disassembly of Dps-DNA complexes and proteolysis of Dps when the cells enter exponential phase (Schmidt et al. 2009). Thus, Dps is a substrate for the N-end rule degradation pathway, where the half-life of a protein is determined by the nature of its N-terminal residue.

Non-specific DNA Binding Property

DNA binding proteins found in cells have various interacting modes with DNA and can be broadly classified as sequence specific and non-sequence specific interactions. Sequence specific DNA protein interactions involve special structural motifs on a protein which is recognised by a consensus sequence on the DNA, e.g. restriction endonucleases recognising palindromic sequences on DNA. On the other hand, non-specific DNA protein interactions predominantly involve ionic interactions with the sugar phosphate backbone of DNA. Some of these non-specific DNA binding proteins gives rise to large supramolecular structures with their association with DNA, e.g. as seen in the case of genome organization by histones in eukaryotes. Also, binding of multiple proteins to long DNA strands allows condensation and makes DNA inaccessible to other environmental factors, giving physical protection (Ganguly et al. 2012).

Studies have shown that when Dps interacts with DNA, highly stable crystals are formed within which DNA is sequestered and protected against various conditions of stress. This has been shown under in vitro and in vivo conditions (Wolf et al. 1999). DNA-Dps co-crystallization presents a binding mode that provides wide range protection by non-specific DNA sequestration. Although the crystalline state is incompatible with life, sequestration of important biological macromolecules in crystalline assemblies may provide adequate protection under conditions of stress. Such ordered Dps-DNA co-crystals have a condensed organisation, which also dramatically enhances the stability of DNA and can endure prolonged starvation and other adverse conditions (Frenkiel-Krispin et al. 2001).

This section explores the mechanism of non-sequence specific DNA binding property of Dps, which has been studied by Atomic force microscopy, X-ray crystallography and more recently Cryo electron microscopy.

Studies on Dps Deletion Mutants

Dps does not have any known DNA binding motifs. In this scenario, initial studies on Dps-DNA binding were carried out by looking at how deletions affect binding to DNA. In *E. coli*, to establish the effect of the electrostatic interactions between the positively charged residues on the surface of the N-terminus and negatively charged DNA, systematic deletions of the lysines at the N-terminus were carried out. Dps with a deletion of two lysines at the N-terminus was able to bind DNA but could not carry out DNA condensation, whereas a deletion of all the three lysines rendered it not able to bind DNA. The property of Dps to self-aggregate was implicated in its ability to condense DNA by protein–protein aggregation (Ceci et al. 2004) (Fig. 3.4). A long flexible N-terminus (Fig. 3.5a) with charged residues like lysines and arginines are seen in a number of Dps homologs, pointing to similar mechanisms of DNA binding. This was demonstrated by Dps homologs in *D. radiodurans*, where deletion of the N-terminus abrogated DNA binding (Bhattacharyya and Grove 2007).

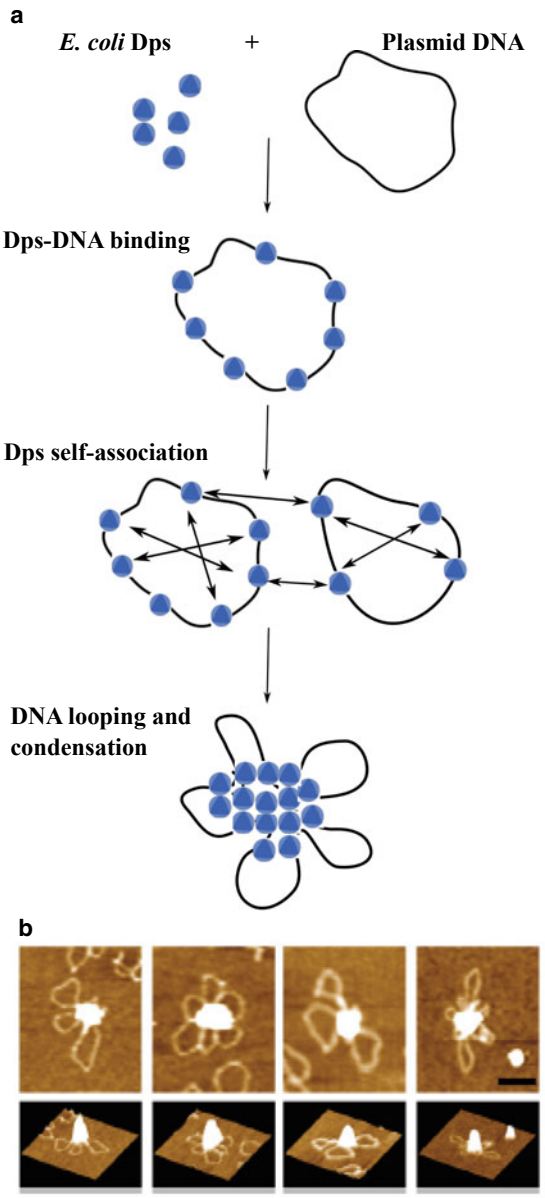


Fig. 3.4 *DNA condensation through E. coli Dps self-aggregation* **a** *Model proposed by Ceci* et al. *for DNA condensation.* Dps is represented as blue circles. The binding of Dps to DNA is promoted by lysines at the N terminus of Dps. **b** *AFM images of Dps-DNA complexes.* Upper panel shows top view and lower panel shows the corresponding lateral 3D views. Scale bar is 100 nm. Figure adapted with permission from Nucleic Acids Res. 2004. 32(19): 5935–5944

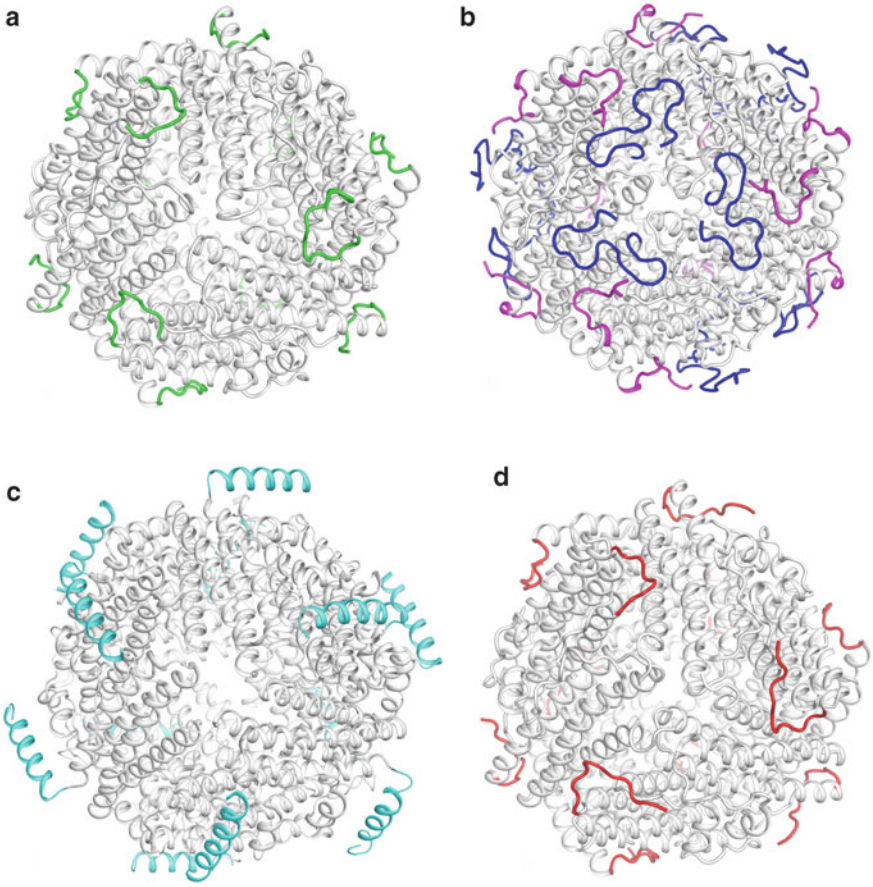


Fig. 3.5 *DNA binding regions of Dps homologs from different species.* **a** *E. coli* Dps dodecamer (PDB ID: 1DPS) with DNA binding flexible N-terminal residues 9–21 represented in green, residues 1–8 are missing in the crystal structure due to flexibility. **b** *M. smegmatis* Dps1 dodecamer (PDB ID: 1VEI) with the interlocking N-termini (magenta) and C-termini (blue). **c** *L. lactis* Dps (PDB ID: 1ZUJ) dodecamer has a helical ordered N-terminus (cyan). **d** *A. tumifaciens* Dps dodecamer (PDB ID: 1O9R) N-terminal region in red is immobilized on the protein surface and unable to bind DNA

A variation is the *M. smegmatis* Dps protein MsDps1 which seemed to bind DNA through a mobile and positively charged C-terminus with 3 lysines and 2 arginines, and a deletion of the 16 C-terminal residues resulted in a loss of DNA binding activity (Roy et al. 2007). Deletion of the entire C-terminal tail resulted in an open decamer assembly, indicating that the C-terminal tail has a role in DNA binding and oligomerisation (Fig. 3.5b). An ordered alpha helical N-teminus in *L. lactis* Dps can condense only long stretches of DNA > 4000 bp. The exact mechanism for this is unclear, but the authors proposed that the decreased flexibility of the N-terminus due to its helical conformation could have a role in the size limitation of DNA binding

(Stillman et al. 2005) (Fig. 3.5c). The *H. plylori* Dps, HP-NAP has a positively charged protein surface at neutral pH and binds DNA through the protonated amino acids on the protein surface (Chiancone and Ceci 2010; Ceci et al. 2007). Dps from *T. erythraeum* lacking a positively charged N-terminus and a truncated C-terminus, could use a mechanism similar to HP-NAP to bind DNA (Castruita et al. 2006).

Some homologs of Dps like the *L. innocua* Dps, *B. anthracis* Dlp-1 and Dlp-2, *C. jejuni* etc are characterised by a short N-terminus and are unable to interact with DNA (Bozzi et al. 1997; Papinutto et al. 2002; Ishikawa et al. 2003). *A. tumifaciens* Dps N-terminal region is immobilized on the protein surface and is not accessible for DNA binding (Fig. 3.5d) (Ceci et al. 2003). Also, pH and salt compositions are shown to have an influence in the ability of Dps to bind and/or condense DNA in vitro, so these factors have to be taken into consideration for comparison of studies across different Dps homologs.

Higher Order Dps-DNA Assemblies Visualised Through Atomic Force Microscopy

Dps condenses DNA by self-association giving rise to large Dps-DNA aggregates which do not enter agarose gels. This was first demonstrated in *E. coli*, where Dps binds DNA through its charged lysine residues. Further through self-aggregation it brings about the looping and condensation of DNA (Fig. 3.4). The interaction to DNA is directed towards condensation only at pH values where the lysine residues are fully protonated. But *H. pylori* Dps HP-NAP uses a different mechanism, as it is not able to self-aggregate, but can bind and condense DNA at slightly acidic pH values where the histidines which mediates DNA binding are protonated (Ceci et al. 2007).

The Dps from *M. smegmatis* MsDps1 when overexpressed in cells, promoted a toroidal structure of Dps-DNA. Whereas, with the overexpression of the second *M. smegmatis* Dps MsDps2 a coral reef structure was formed (Fig. 3.6a) (Ghatak et al. 2011). *D. radiodurans* Dps2-DNA interaction was shown to form toroidal structures with DNA, similar to the arrangement in MsDps1 (Fig. 3.6b) (Santos et al. 2015). The higher-order assemblies of Dps-DNA shown by AFM and EM are a result of DNA condensation or co-crystallization. Thus, looping and condensation of DNA is seen when Dps is added to linear or circular stretches of DNA. Toroids, coral reef structures and co-crystals are formed within bacterial nucleoids under very late stationary phase or when the protein is overexpressed in cells. Additionally, co-crystals with DNA have been shown to form under in vitro conditions.

It was shown that the effect of Dps on the nucleoid state can vary depending on different stages of growth or on the species of bacteria. The overexpression of Dps in *E. coli* in the log phase induces neither nucleoid condensation nor DNA-Dps co-crystallisation, whereas in the stationary phase it results in nucleoid condensation. On the other hand, the *S. aureus* Dps homolog mrgA (Morikawa et al. 2006) causes

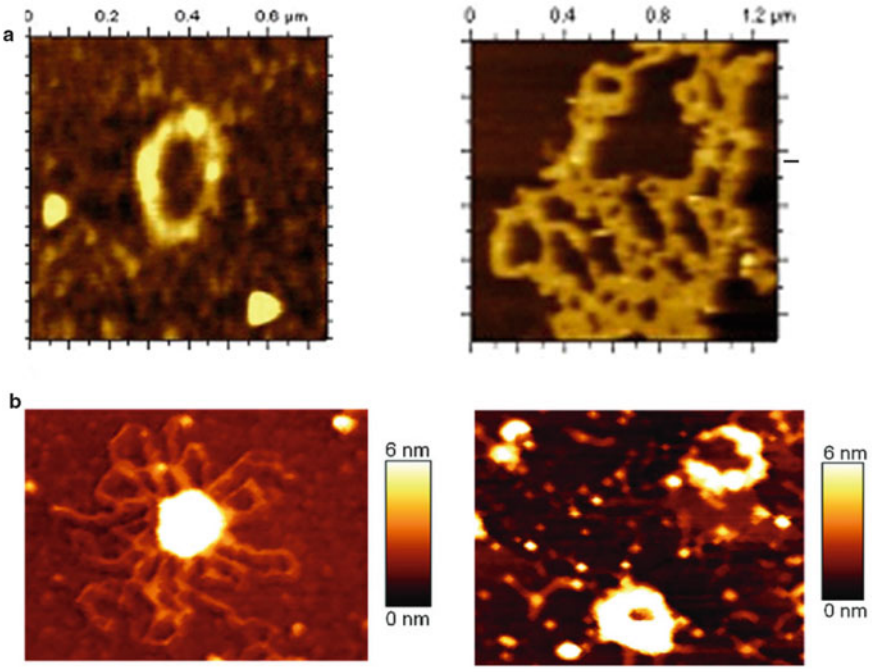


Fig. 3.6 *Higher order Dps-DNA assemblies visualised by Atomic Force Microscopy.* **a** The lysate of *M. smegmatis* mc^2155 overexpressing Dps1 at late stationary phase, shows toroidal assemblies of Dps-nucleoid with a diameter of 70–100 nm (left panel). Late stationary phase cell lysate of mc^2155 overexpressing Dps2 shows coral reef structures with nucleoid (right panel). (Images reproduced from PLOS ONE. 2011. 6(1): e16019) **b** *D. radiodurans* Dps1-DNA complex at pH 6.5 shows DNA condensation (left panel) and DrDps2-DNA toroidal assemblies at pH 6.5 (right panel). Color scale represents the height of *D. radiodurans* Dps-DNA complex. Images reproduced with permission from FEBS J, 282: 4307–4327

nucleoid condensation in both log and stationary phases on over-expression in cells. This was due to the action of Fis, a bacterial nucleoid associated protein abundant in the log phase which inhibits nucleoid condensation by Dps during this stage of growth. Also, Fis is present in gamma Proteobacteria which includes *E. coli* but not *S. aureus*, which explains why there is no regulation of DNA condensation in the log phase when Dps is overexpressed in the latter organism. Thus, the organisation of Dps into higher order assemblies with DNA is species-specific and also is an interplay of Dps and other nucleoid binding proteins in cells during different stages of growth.

DNA-Dps Biocrystals

Living systems when exposed to severe environmental stresses, may sequester vital macromolecules in intracellular crystalline assemblies for efficient protection. Under these conditions often defense strategies based on enzyme catalysis can no longer work efficiently due to energy restrictions, so DNA protection is often achieved through physical transactions like phase separation and phase transitions (Wolf et al. 1999). A defense strategy involving co-crystallization of DNA and Dps was observed under in vitro and in vivo conditions in *E. coli* by Wolf and colleagues. They showed that purified Dps and DNA interact and almost instantaneously form stable crystals (Fig. 3.7a) within which DNA is sequestered and protected from possible environmental assaults. Similar crystalline structures are also formed in vivo in cells where Dps is overexpressed or in wild-type bacteria under starvation conditions (Fig. 3.7b).

Recently, two different types of Dps-DNA co-crystals were observed by cryo-electron tomography (Kamyshinsky et al. 2019). Subtomogram averaging of first type of co-crystals revealed a triclinic crystal lattice (Fig. 3.8a). The second type of co-crystals adopted a cubic crystal lattice (Fig. 3.8b), where each Dps is surrounded by six DNA strands, unlike in triclinic co-crystals where each Dps is surrounded by four DNA molecules. But the exact nature of Dps-DNA interactions could not be revealed in the low-resolution reconstructions and future single particle cryo-EM studies on individual DNA-Dps complexes or 2D crystals may achieve this.

Modes of DNA Protection and Regulation

The multifarious protection of DNA by Dps is by virtue of three intrinsic properties of the protein: DNA binding, iron sequestration and ferroxidation. Moreover, Dps is proposed to have roles in gene regulation during starvation, by controlling expression at the transcriptional level. This section explores these claims in the light of recent findings, as well as summarising already existing knowledge of DNA protection mechanisms.

Protecting DNA From Damage

Dps prevents DNA damage by direct physical sequestration and indirectly through its ferroxidation activity. These biochemically separate activities function jointly to preserve DNA integrity (Fig. 3.9) (Calhoun and Kwon 2011). Through its DNA binding ability, Dps was shown to confer protection from UV and gamma radiation damage, which induces potentially harmful mutations in DNA (Nair and Finkel 2004). Another instance of Dps endowing resistance to cells enduring acidic conditions is through acid tolerance in *E. coli* which helps the bacterial passage through

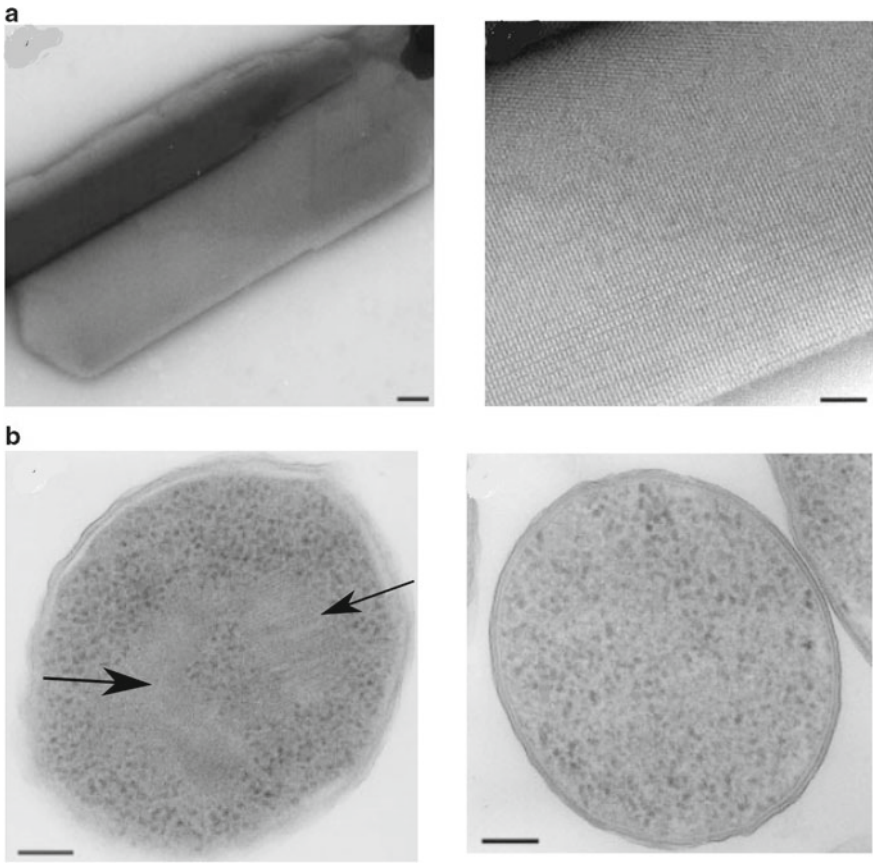


Fig. 3.7 *Electron microscopy images of Dps-DNA co-crystals* **a** Dps incubated for 15 s with closed circular DNA in vitro (left panel), scale bar 100 nm and its higher magnification, scale bar 40 nm (right panel). **b** Crystalline assemblies of Dps-DNA in a starved WT *E. coli* cell indicated by black arrows (left panel) and dps knockout cell (right panel), scale bars are 100 nm. Images reproduced with permission from Nature 1999. 400: 83–85

gastric barrier of humans. It was shown that acid stress led to damage of chromosomal DNA, accentuated in *dps* mutants (Jeong et al. 2008), showing a role for Dps in shielding DNA against acid stress. One of the earliest mechanisms which explains Dps protection of DNA against H_2O_2 mediated oxidative stress was found to be by direct binding of Dps to DNA, reproduced in in vitro systems (Martinez and Kolter 1997). Dps null mutants were also more sensitive to some types of metal toxicity, such as toxicity from copper and iron during stationary phase. Temperature dependent adaptation was also seen to a less extent in these mutants as compared to WT cells, during thermal stress (Nair and Finkel 2004). Dps expression is also upregulated by the OxyR transcriptional regulator, in response to oxidative stress (Altuvia et al. 1994).

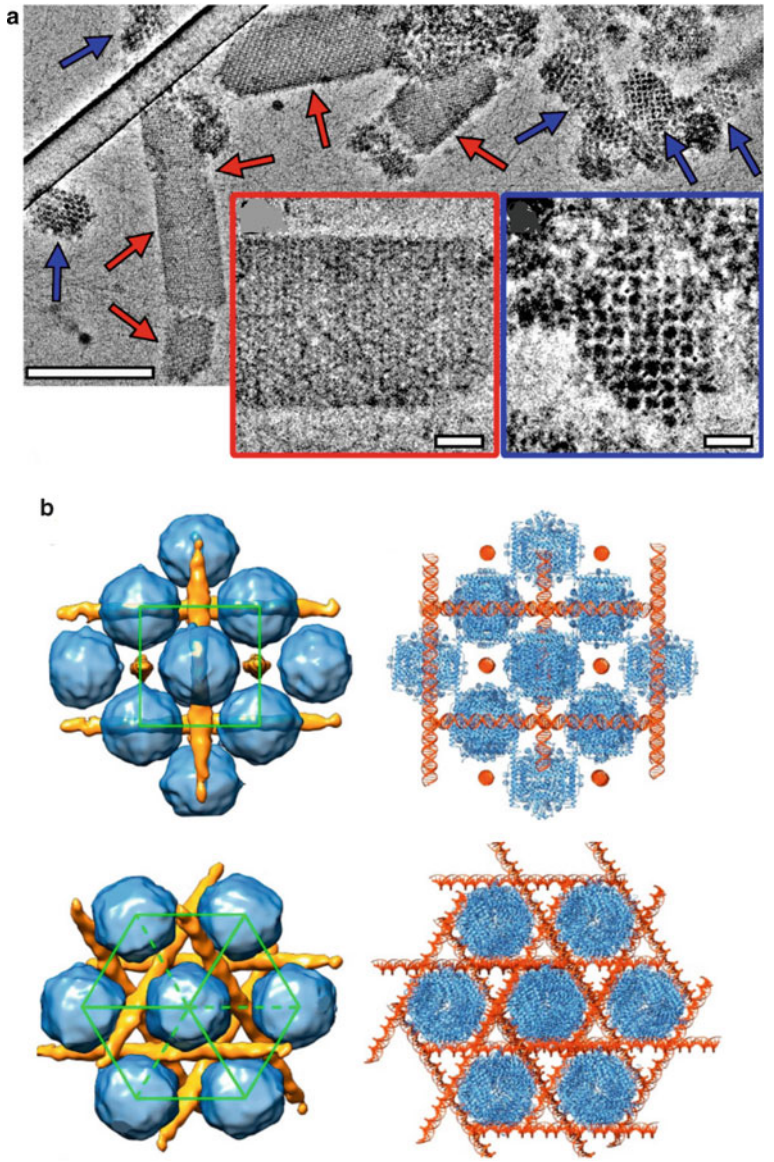


Fig. 3.8 *Dps-DNA co-crystals* **a** CryoTEM images of Dps-DNA co-crystals, red arrows indicate ones with a triclinic crystal lattice and the red inset is a high magnification of the same. Blue arrows indicate co-crystals with a cubic crystal lattice, a higher magnification is shown in the inset with blue outline. Scale bars are 200 nm and inset 20 nm. **b** 3-D reconstructions based on subtomogram averaging of cubic Dps-DNA co-crystals, in two different views. DNA is in orange and Dps are blue spheres with unit cell shown in green (left). Corresponding models of Dps and DNA fitted in the EM densities are shown in the right panel. Images reproduced from Biomolecules. 2020.10: 39

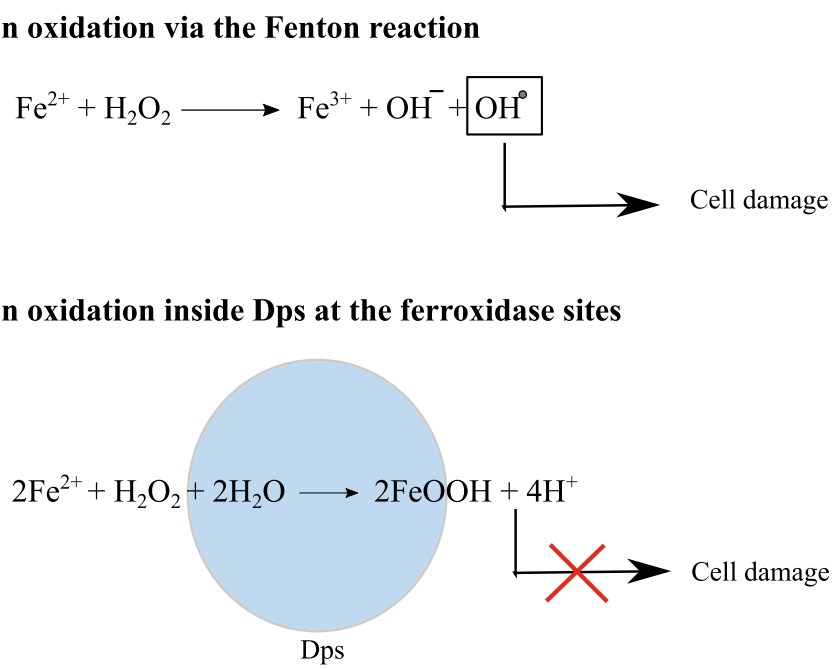


Fig. 3.9 Oxidation of Fe(II) occurs via Fenton reaction releasing hydroxyl free radicals, which causes cell damage (top panel). Oxidation of Fe(II) happening inside the Dps shell at ferroxidation sites proceeds without production of free radicals thereby protecting cells from oxidative damage (bottom panel). Figure adapted with permission from Journal of Applied Microbiology. 2010. 110: 375–386

Dps can also protect the genome from a distance using DNA charge transport through the DNA base pair π-stack. This was demonstrated in *E. coli* Dps, where an intercalating ruthenium photooxidant was used to generate localised DNA damage to guanine repeats, which are the sites of lowest oxidation potential in DNA. When iron loaded Dps was titrated against this modified DNA as opposed to apo-Dps, it significantly reduced oxidative damage of DNA. The authors propose that the iron loaded Dps is oxidised by DNA mediated oxidation through charge transport, as there was no direct contact between the ruthenium photooxidant and Dps (Fig. 3.10) (Arnold and Barton 2013). Thus through DNA charge transport, Dps could respond to an oxidative affront happening around a 100 base-pairs away.

Conserved aromatic residues like tyrosine and tryptophan near the ferroxidase site, can act as electron transfer intermediaries, allowing for rapid transfer of electrons between the ferroxidase centre and outside of the protein shell. In *L. innocua* Dps, a double mutant lacking both the aromatic residues near the ferroxidase site showed attenuation of DNA protection and the DNA was degraded by hydroxyl radicals generated through ferrous oxidation by H_2O_2. DNA protection assays indicated that the presence of aromatic residues limited the release of hydroxyl radicals into the solution and subsequent oxidative damage of DNA. Thus the aromatic amino acids

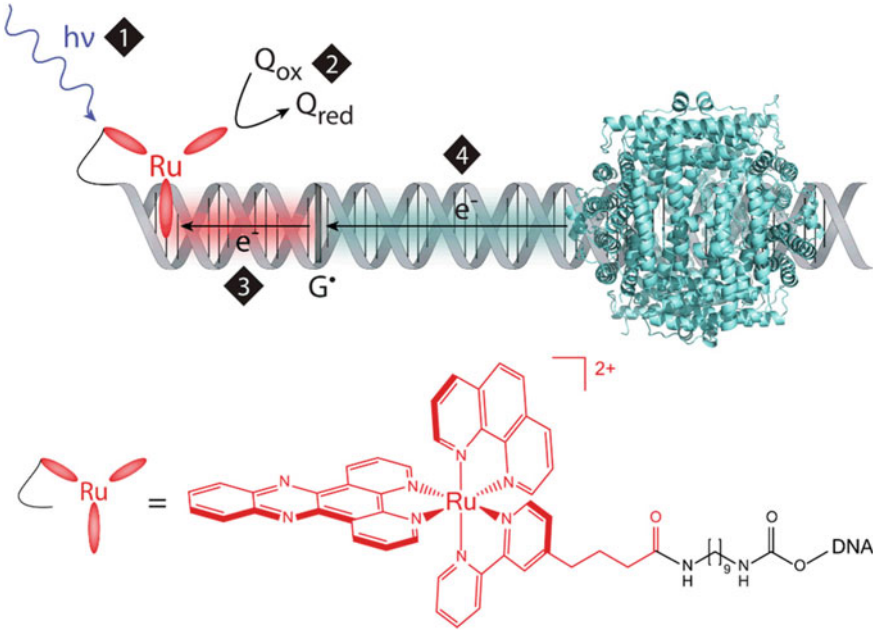


Fig. 3.10 *Schematic of DNA protection through charge transport*. Steps: 1. A DNA intercalated ruthenium(II) photooxidant is excited by visible light. 2. Ru(II) is oxidatively quenched to Ru(III) by a diffusing quencher (Q). 3. Ru(III) abstracts an electron from DNA and the electron hole equilibrates along the π-stack and localizes to the most easily oxidised base like guanine (G•). 4. DNA charge transport from Dps to guanine radical to fill the electron hole is a possible long distance protection mechanism. Lower panel shows the structure of Ru(II) covalently tethered to DNA via diaminononane linkage. Figure reproduced with permission from J Am Chem Soc. 2013.135(42):15,726–15,729

near the ferroxidase centre could act as a trap for electron holes generated by oxidation of ferrous iron by H_2O_2 (Bellapadrona et al. 2010). Also, these residues could play a role in DNA mediated oxidation of Dps through charge transport. These findings further highlight the interplay of ferroxidation and DNA binding property for DNA protection.

Nucleoid Condensation in Response to Stress

The genome in bacteria is folded into a compact structure called the nucleoid. In order to reduce the volume to fit inside a cell, nucleoid associated proteins or NAPs bind and condense DNA by bridging, bending or wrapping (Fig. 3.11a–f). In addition, other factors like molecular crowding and DNA supercoiling contribute to further compaction of DNA. DNA bending NAPs like IHF, HU and Fis act together with DNA bridging proteins like H-NS or Lrp, with expression varying according to the

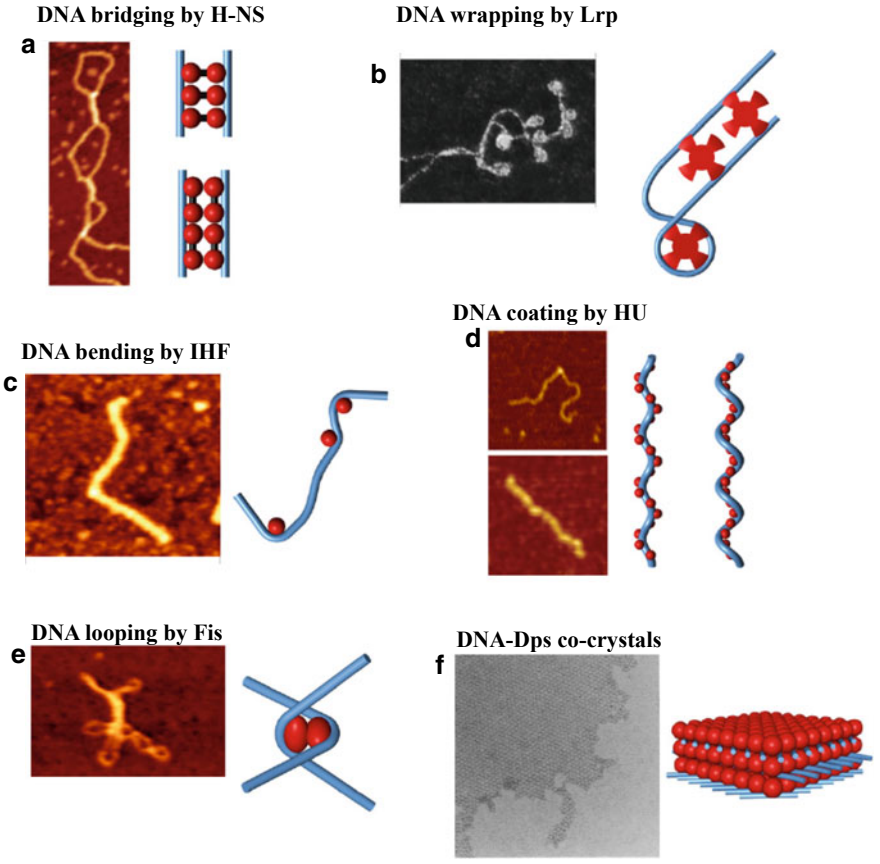


Fig. 3.11 *Architectural roles of common nucleoid associated proteins*. Images on the left show microscopy images of DNA with the respective NAP, and the cartoon representation is shown on the right. Images reproduced with permission from J. Struct. Biol. 2006. 156: 262–272 and Genes & Dev. 1992. 6: 2646–2654

growth phases of the cells. Fis has optimal expression during exponential phase (1 Fis/460 bp), after which HU is expressed (1 HU/550 bp). In early stationary phase IHF reaches its expression peak (1 IHF/335 bp). Dps is the most abundant protein in late stationary phase (1 Dps dodecamer/300 bp) (Luijsterburg et al. 2006).

During fast exponential phase of growth with high transcriptional activity, the nucleoid is associated with clustered transcriptional foci. In stringent response where inhibition of transcription happens, the nucleoid decondenses to form an elongated structure. In early and stationary phase ring-shaped toroids containing DNA and Dps and eventually the Dps-DNA crystalline lattice appears (Fig. 3.12), where the DNA is transcriptionally inactive (Travers and Muskhelishvili 2005). To summarise the changes in the expression of NAPs, in the exponential growth phase Fis is the most abundant protein with the following order of abundance: Fis > Hfq > HU > StpA

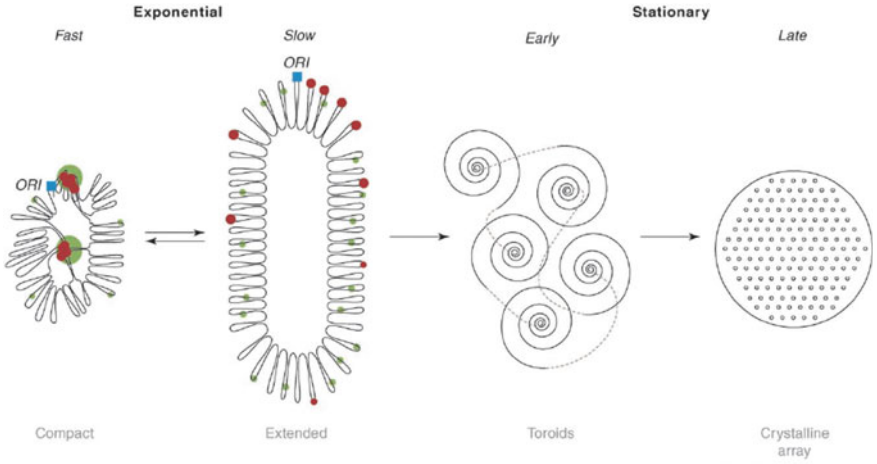


Fig. 3.12 Alterations in the nucleoid organisation as a function of growth phase in a bacterial cell. Figure reproduced with permission from Current Opinion in Genetics & Development. 2005. 15:507–514

> H-NS > CbpB > Dps. In early stationary phase, Fis disappears and the order of abundance changes to: Dps > IHF > Hfq > HU > CbpA > StpA (Ali Azam et al. 1999).

Role in Transcription Regulation

The above sections have described the mechanisms by which multifaceted Dps proteins protect cells against stresses through DNA binding, iron binding and sequestration and ferroxidation. Another possible mode of protection is by the regulation of gene expression by modulating the structure of DNA thereby controlling access to the transcription machinery. Two dimensional gel electrophoresis of cell lysates from WT and *dps* null mutant shows differential expression of proteins (Almiron et al. 1992). The compacted Dps-DNA structures, along with differences in protein expression patterns suggested a role for Dps as a pleiotropic regulator of transcription when cells enter the stationary phase.

But, a recent study has shown that *dps* deletion causes no significant change in the gobal transcriptional patterns and only mild alterations in the proteome, in vivo during the stationary phase. Dps had no effect on RNA polymerase (RNAP) initiation at physiologically relevant concentrations. Single molecule transcription assays used in vitro to probe effects of Dps-mediated DNA compaction revealed that Dps binding did not induce RNAP pausing or arrest transcriptional elongation of DNA (Janissen et al. 2018). These findings suggest that Dps is a DNA binding protein that completely decouples DNA condensation from transcriptional regulation, in contrast to histones

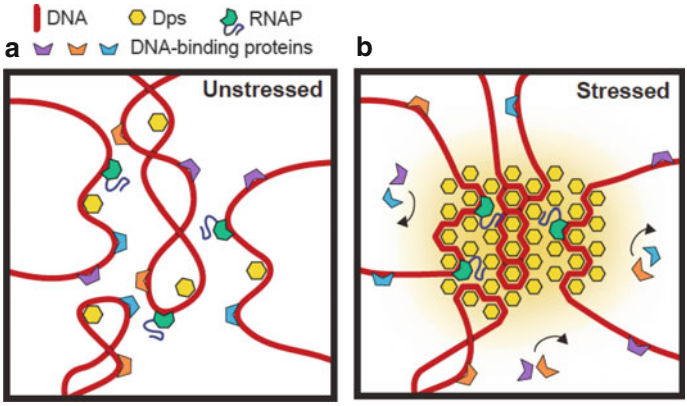


Fig. 3.13 *Model of DNA Protection by Dps* **a** Under normal conditions in cells, *E. coli* Dps binds DNA but is unable to condense the nucleoid. **b** Under conditions of stress, Dps condenses most sections of the nucleoid, creating phase-separated organelles, which is accessible to RNA polymerase but inaccessible to other nucleoid binding proteins. Figure reproduced with permission from Cell. 2018. 174(5):1188–1199

and other bacterial NAPs like HU and H-NS. This could provide bacteria greater freedom to tailor transcriptional responses to different stresses at the same time as protecting the genome from damage.

Dps shows co-operativity in binding to DNA (Ganguly et al. 2012) and this requires multiple nearest neighbour interactions between Dps dodecamers to stabilize the relatively weak interactions with DNA. Thus, Dps has a cumulative affinity for DNA, despite having low affinity for individual contacts on DNA (Vtyurina et al. 2016). Therefore, proteins such as RNAP which establishes stable interactions with DNA could displace Dps. But it does not explain how some others like the restriction enzymes with high affinity for DNA could be excluded from acting on Dps-DNA complexes. The dynamic behaviour of Dps-DNA binding could indicate that it forms a phase-separated organelle with DNA, permeable to RNAP. These evidences suggest that the ability of Dps to protect cells during stress is a direct effect of DNA binding rather than activation or repression of specific genes. This dynamic nature of Dps assembly on DNA also ensures that transcription can continue even under extreme conditions of stress (Fig. 3.13) (Janissen et al. 2018).

Ferroxidation Properties

Iron plays a vital role in the processes of living organisms and the acquisition and storage of metal in a non-toxic and bioavailable form is a primary requirement for cells. Iron starvation is one of the major obstacles for successfully establishing growth. Proteins of the ferritin family, like ferritins, bacterioferritins and Dps are

the major iron storage entities in cells, with Dps found only in prokaryotes. Dps are dodecameric nano-compartments having roughly spherical structures with a hollow core for storing iron. Twelve identical monomers assemble through a 23 symmetry to form 12-mers, and are frequently referred to as miniferritins, smaller in size than the 24-meric ferritins. They can store ~500 iron atoms per dodecamer and can oxidise iron to its inert ferric form (Bozzi et al. 1997). Studies that examine the mechanism of iron uptake, ferroxidation and release in response to the needs of cells, has been carried out through mutational and structural analyses. This section summarises the molecular mechanisms of how Dps processes iron. Being iron storehouses in cells, Dps proteins share considerable structural similarity to ferritins. An evolutionary link to ferritins in the light of recent studies showing a structural basis of a switch between Dps and ferritins is also discussed.

Iron Uptake and Release

To better understand how Dps sequesters iron, it is necessary to examine the structure of these dodecamers. Each monomer is an alpha helical bundle, with four helices named A–D, and a short BC helix in between B and C helices (Fig. 3.2a). These monomers have a tendency to form dimers, which are thought to be the first intermediate of oligomerisation. The ferroxidase center is placed in the dimeric interface in between two subunits. Due to its symmetrical arrangement, Dps exhibits two type of three-fold interfaces, namely the ferritin-like three-fold and the dps-like three-fold interfaces. The ferritin-like three-fold interface is important in the channelling of iron into the protein, whereas the dps-like interface is responsible for the protein folding into a dodecamer. A model for iron incorporation can be described in three steps: (1) Fe(II) atoms enter the protein cavity through the four ferritin-like three fold interfaces of the dodecamer. (2) Fe(II) atoms are oxidised at the ferroxidase centre. (3) Oxidised Fe(III) atoms move towards nucleation sites in the interior of the cavity. (4) Fe(III) mineral core is formed, with the possibility of further incoming Fe(II) atoms getting oxidised on the surface of this growing mineral core (Fig. 3.14) (Pulliainen et al. 2005).

Like ferritins, Dps has a roughly spherical structure with a hollow core and pores at the ferritin-like interface, making them highly suited for iron sequestration and storage. The ferritin-like interface in Dps forms a funnel-shaped channel with a wide mouth facing the solvent side and which narrows towards the interior of the protein cavity (Fig. 3.15a, b). This channel is lined by hydrophilic residues, mainly negatively charged aspartate and glutamate residues, which trap and internalize iron into the cavity (Williams et al. 2014). A number of studies in ferritins point to the ferritin-like pore as the channel of entry and exit. Mutational analyses in *S. suis* and *L. innocua* Dps indicated the importance of the conserved aspartate residues in this channel, whereby any substitutions at these loci led to changes in the rate of iron uptake and release (Pulliainen et al. 2005; Bellapadrona et al. 2009).

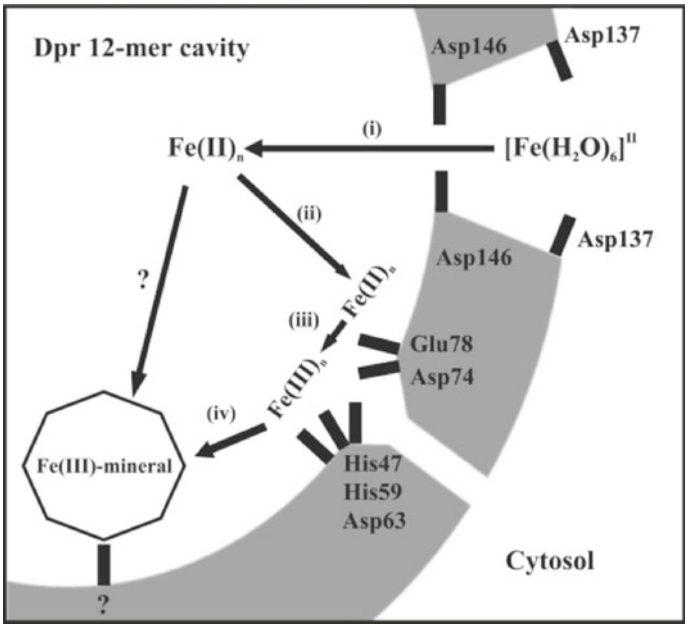


Fig. 3.14 *Schematic of iron oxidation and storage by Dps proteins, with S. suis Dpr as a representative example.* Steps: (i) Fe(II) enters the protein cavity through the four hydrophilic pores each within the ferritin-like three-fold interfaces, possibly attracted by negatively charged residues like aspartates which lines the pore. (ii) Fe(II) gets oxidised at the twelve inter-subunit ferroxidation sites, each located between two subunits related by a two-fold interface (iii) and (iv) Fe(III) ions move to nucleation sites where Fe(III) mineral formation is initiated, and the mineral deposition continues to grow with more Fe(II) getting oxidised. "?" represents a path for Fe(II) ions to be oxidised directly on the growing mineral core, which could happen at a faster rate and explains the sigmoidal nature of the ferroxidation curve in Dps proteins. Figure reproduced with permission from Mol Microbiol. 2005. 57(4): 1086–1100

A mechanism of iron uptake and release was demonstrated on the basis structural evidence in *M. smegmatis* Dps, MsDps2. An assembly of oppositely charged histidine-aspartate cluster at the ferritin-like interface was identified, which guards the narrowest point of the channel. Site specific variants which disrupted this ionic knot, showed a defective mechanism of iron uptake and release, highlighting a gating mechanism that happens by alterations in the side chains of the residues lining the channel and does not change the over-all stability of the protein. Iron release from iron loaded Dps can be carried out by a combination of reductants and chelators. Reducing agents like sodium dithionite converts Fe(III) to Fe(II) which is chelated by agents such as 2,2′-bipyridyl to form Fe(2,2′-bipyridyl)$_3^{2+}$ complex and forms the basis of iron release assays in vitro. The kinetics of iron release exhibits an early fast rate which corresponds to the release of iron bound near the pores and easily accessible to chelators, followed by a slower rate which coincides with the dissolution of iron oxide in the protein core. Replacing the aspartates at the narrow opening with

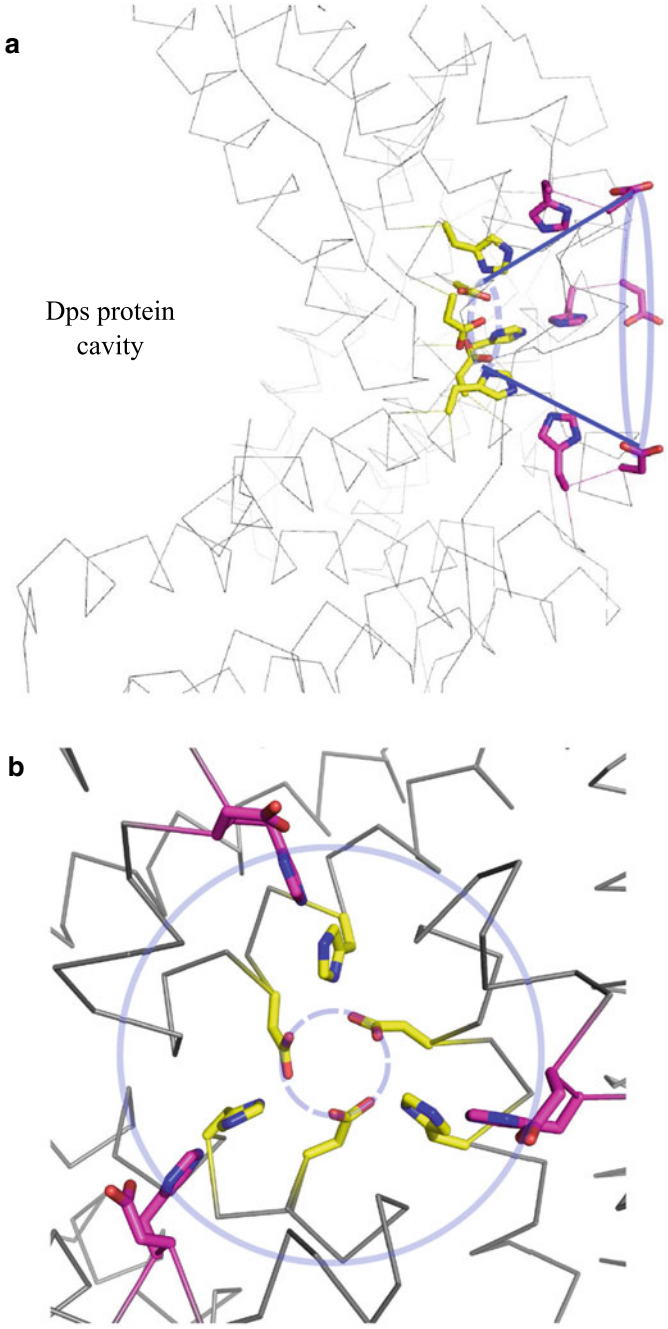


Fig. 3.15 *Funnel shaped channel for iron entry in Dps* **a** Side view and **b** Surface view with a continuous blue circle revealing the diameter of the mouth of the funnel-shaped channel and dashed circle shows the narrowest part of the channel. Residues important for iron channelling are shown as sticks (PDB ID: 2Z90)

bulkier histidine residues reduce the rate of iron release by several fold (Williams et al. 2014).

However, the molecular mechanisms of iron release in vivo to meet iron requirement in cells during conditions of iron demand, is less understood. In vivo iron release studies in ferritins have demonstrated the cytosolic release of iron by ferritin degradation within lysosomes (Linder 2013). It is thought that once the ferritins are degraded in the lysosomes, the exposed ferrihydrite core maybe reduced/chelated and transported to the cytosol. This mechanism may not translate to Dps, as Dps is present only in prokaryotes. The N-end rule degradation pathway of Dps via ClpX/ClpP could release the iron deposited within, making it accessible to chelators and other iron transporters in cells.

Conserved Ferroxidase Active Site

The first structural evidence of a ferroxidase site was revealed in the X-ray crystal structure of *L. innocua* Dps (Ilari et al. 2000). This first ferroxidase centre identified contained only one iron atom, contrary to the conventional di-iron centres found in ferritins. Since then di-iron FOCs have been identified in other Dps homologs, where the second iron displayed a partial occupancy. Unlike the FOCs in ferritins which are contained within the four-helix subunit, the FOC in Dps is between two subunits related by a two-fold. The iron in LiDps FOC is coordinated by carboxylate residues and one or two histidine residues that play a role in the redox process. The two FOCs in a two-fold interface are separated by a distance of 21.5 Å.

A di-iron FOC was indicated in the crystal structure of *B. brevis* Dps, where two neighbouring di-nuclear ferroxidase centers responsible for the oxidation of Fe(II) ions was identified. The iron at site 1 is tightly coordinated to NE2 of His 31 in subunit A, carboxylate group of Asp58 and OE2 of Glu62 from subunit B, where A and B are related by a two-fold symmetry. Iron at site 2 is loosely bound to NE2 of His43 (subunit A), OE1 of Glu62 (subunit B) and OE1 of Glu47 in subunit A via a water molecule (Fig. 3.16). A μ-oxo bridge links the iron atoms of site 1 and 2, representing a reaction intermediate Fe(III)-O-Fe(III) during the oxidation of Fe(II) (Ren et al. 2003). Like most active site residues, the ferroxidation site residues are well conserved among Dps homologs from various species.

Iron Gating Mechanisms

Dps achieves iron channelling by very little changes to its overall conformation, being very stable protein cages. A recent study by our group through co-crystallisation of Dps with iron, showed that flexible aspartates could propel iron from the entry site to the ferroxidation centre. A conserved arginine residue forms a network of interactions which stabilizes the interface between the ferroxidation and iron entry

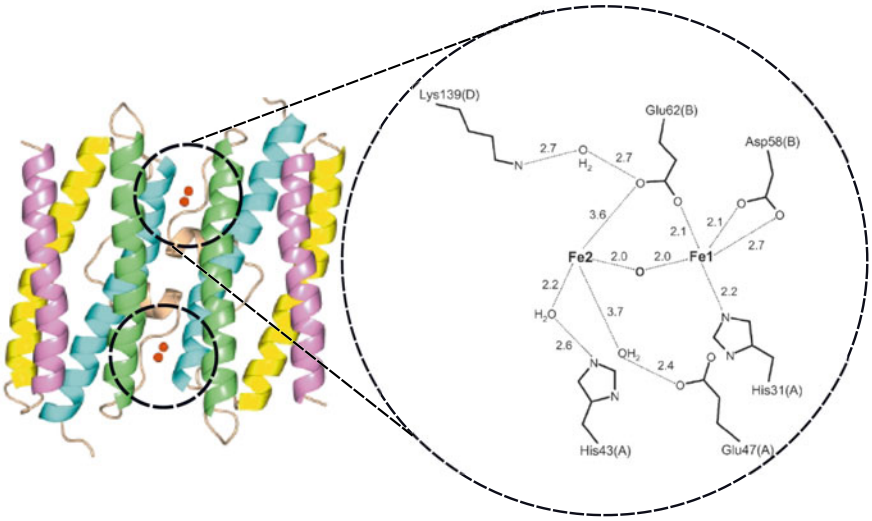


Fig. 3.16 *Dinuclear ferroxidase centre in B. brevis Dps.* Two FOCs are in black discontinuous circles, related by a two-fold interface (PDB ID: 1N1Q). One of the FOCs is magnified on the right panel, showing cordinating residues from subunits A and B. Fe ions at site 1 and site 2 are labelled Fe1 and F2, respectively. Figure adapted with permission from J. Mol. Biol. 2003. 329: 467–477

sites (Fig. 3.17) (Williams and Chatterji 2017). The iron can bind to three sites at the entry channel as shown in Fig. 3.18a. In *M. smegmatis* MsDps2, Asp138 which lines the narrowest region of the pore, shows flexibility which could help in propelling iron to the interior and any substitutions at this site caused a marked reduction in the rate of iron entry. Another, Asp68 at the ferroxidation site was shown to exhibit alternate conformations. Asp68 in its "outward-facing" conformation where the side chain faces away from the iron binding centre, is a possible state when no iron is bound at the site. In an "inward-facing" conformation, the ferroxidation site satisfies the coordination chemistry for iron binding (Fig. 3.18b). A three-tier arrangement of flexible asparates, each tier having three aspartates related by a three-fold interface, lines the path of iron from its entry to the ferroxidation site (Fig. 3.19) (Williams and Chatterji 2017). Further biophysical studies looking at how the movement of aspartate side chains are linked to the passage of iron might throw valuable insights on iron channelling in Dps.

Structural analysis of Dps loaded with iron to saturation levels did not show any appreciable changes in X-ray crystallographic structure. Possibly because an amorphous core of iron cannot be detected by X-rays, no density accounting for iron could be found in any of the co-crystal structures. Co-crystals of varying ratios of iron: dps did not show any appreciable structural variations from low iron-bound forms to higher iron loaded forms. This indicates the extraordinary stability of the Dps cages which can take iron to saturation levels without major conformational adjustments. These proteins could allow iron entry by fluctuations in side chains of the hydrophilic residues lining the pore gates (Williams and Chatterji 2017).

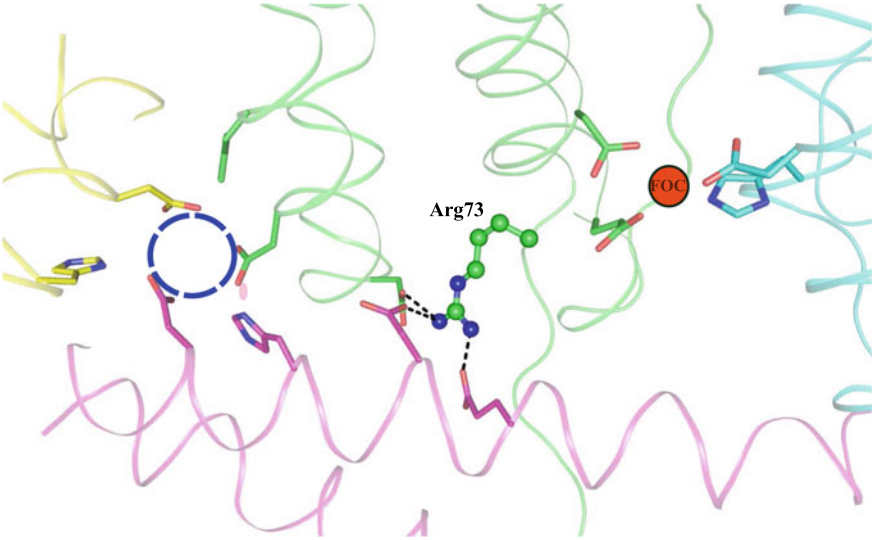


Fig. 3.17 View from the inside of a Dps dodecamer showing conserved arginine stabilizing interactions between the ferroxidation centre (FOC) shown by a red circle and iron entry channel indicated by a blue dashed circle. Each monomer and its residues forming the interface is colored differently. Arg73 is the conserved arginine in the dps homolog MsDps2 (PDB ID: 2Z90)

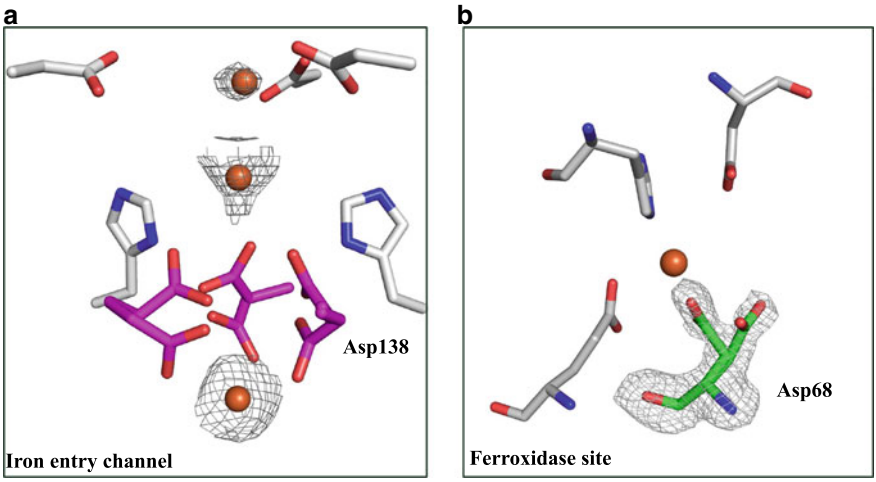


Fig. 3.18 *Alternate conformations of aspartates at* **a** Iron entry channel and **b** Ferroxidase site. Asp68 and Asp138 are the residues in the FOC and iron entry pore, respectively in MsDps2 (PDB ID: 5WW5)

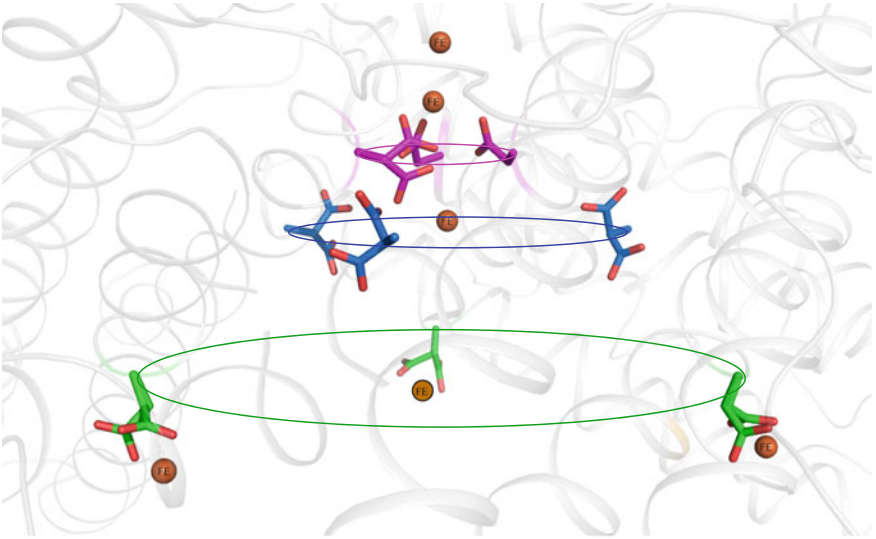


Fig. 3.19 *Three tiers of flexible aspartates help to propel iron to the ferroxidase site (FOC).* Iron entry site residues are in magenta sticks, FOC residues are in green. The aspartate residues in blue are in a site between the active site and entry site, could help the movement of iron between these sites. Fe ions are shown as orange spheres. The three aspartates of each tier are related by a trimeric interface (PBB ID: 5WW5)

Interestingly, small iron oxide clusters have been reported in *H. salinarum* DpsA iron loaded protein crystals (Zeth et al. 2004). A greater mobilization of iron could be required in a hypersaline environment faced by this organism, where iron reduction and release could occur faster in small clusters having higher surface to volume ratios, compared to large iron deposits. As such clusters have not been observed in other Dps homologs, this could be an adaptation strategy unique to this organism in surviving extreme conditions.

Binding to Other Metals

Metal ions like terbium and zinc were shown to inhibit iron incorporation into mammalian ferritins. *Listeria innocua* Dps has also been found to be weakly inhibited by Zn(II) and strongly be Tb(III) (Stefanini et al. 1999). *S. mutans* Dpr could incorporate 11.2 zinc atoms per dodecamer in addition to its capacity to incorporate up to 480 iron atoms (Yamamoto et al. 2002). This led researchers to propose that other metal ions could compete with iron binding at the ferroxidation sites. The structural basis of Zn and Tb mediated inhibition of ferroxidation was shown in *S. suis* where crystal structures with Zn and Tb incorporated Dpr showed the presence of Zn and Tb at the ferroxidase centre and in the same location where iron normally binds

(Havukainen et al. 2008). In vitro studies also confirmed that both ions abolish iron oxidation in Dps at concentrations >0.2 mM. A di-zinc binding in the ferroxidase site was also reported in *S. pyogene*s Dpr (Haikarainen et al. 2010). Divalent metal cations Cu^{2+}, Ni^{2+}, Co^{2+}, Mn^{2+}, Zn^{2+} and Mg^{2+} bound to ferroxidase sites *of S. suis* Dpr was characterised by isothermal titration calorimetry and X-ray crystallography. These ions were found coordinated to the ferroxidase site. The exact physiological function of the ferroxidase centre binding other metal ions is not known, but it could have certain indications on how Dps confers resistance against metal stresses of zinc or copper (Haikarainen et al. 2011).

In *S. solfataricus*, a cellular environment where availability of Mn is higher than Fe, SsDps (*S. solfataricus* Dps) could load Mn and inhibit mineralisation of Fe (Hayden and Hendrich 2010). In competition experiments where Mn^{2+} was already bound to the metal sites, Fe^{2+} was unable to displace the bound manganese, thereby inhibiting ferroxidation. *Kineococcus radiotolerans*, a radiation resistant bacterium having a high ratio of Mn:Fe in the cytosol, can survive ionizing radiations. It was discovered that the ferroxidation centre of this organism can have an Mn-Fe composition as opposed to the usual Fe–Fe dinuclear iron bound at the ferroxidase centre. The authors propose that the hetero-binuclear ferroxidase centre with Mn-Fe could have a role in redox cycling and help in combating oxidative stress (Ardini et al. 2013).

The Dps from the cyanobacterium *T. elongatus* was shown to bind two Zn at the ferroxidase site. The FOC in this dps homolog is unique in having two histidines and only one of the bound zinc can be displaced by incoming Fe ions, giving rise to Zn(II)–Fe(III) complexes as indicated by atomic emission spectra (Alaleona et al. 2010). Another cyanobacterial Dps from *N. punctiforme*, NpDps4 has an atypical ferroxidase site similar to TeDpsA with two additional histidines, which also can bind Zinc. This led the authors to classify a new type of FOC called the His-type FOC (Howe et al. 2018). Zinc bound at the ferritin-like three-fold interfaces seen in the case of *H. pylori* HP-NAP, which are iron entry channels in Dps, indicates that other metals could also enter the dodecamer through the usual iron entry sites (Yokoyama and Fujii 2014).

Evolutionary Link with Ferritins

Ferritins are 24-meric proteins with an internal cavity which can store and oxidise iron, like their smaller 12-meric counterparts, Dps. They have a 432 octahedral symmetry with subunits forming four-fold, ferritin-like three-fold and two-fold symmetries. Dps has a 23 tetrahedral symmetry, with subunits related by a ferritin-like three-fold, Dps-like three-fold unique to Dps and two-fold symmetries (Fig. 3.20). The ferritin-like three-fold interface is important for iron channelling, whereas the Dps-like three-fold of Dps and the four-fold interface in ferritins are important for protein folding into higher oligomers. In Dps, the AB loop between helices A and B forms stable interactions at the dps-like trimeric interface, and a mutation of a highly

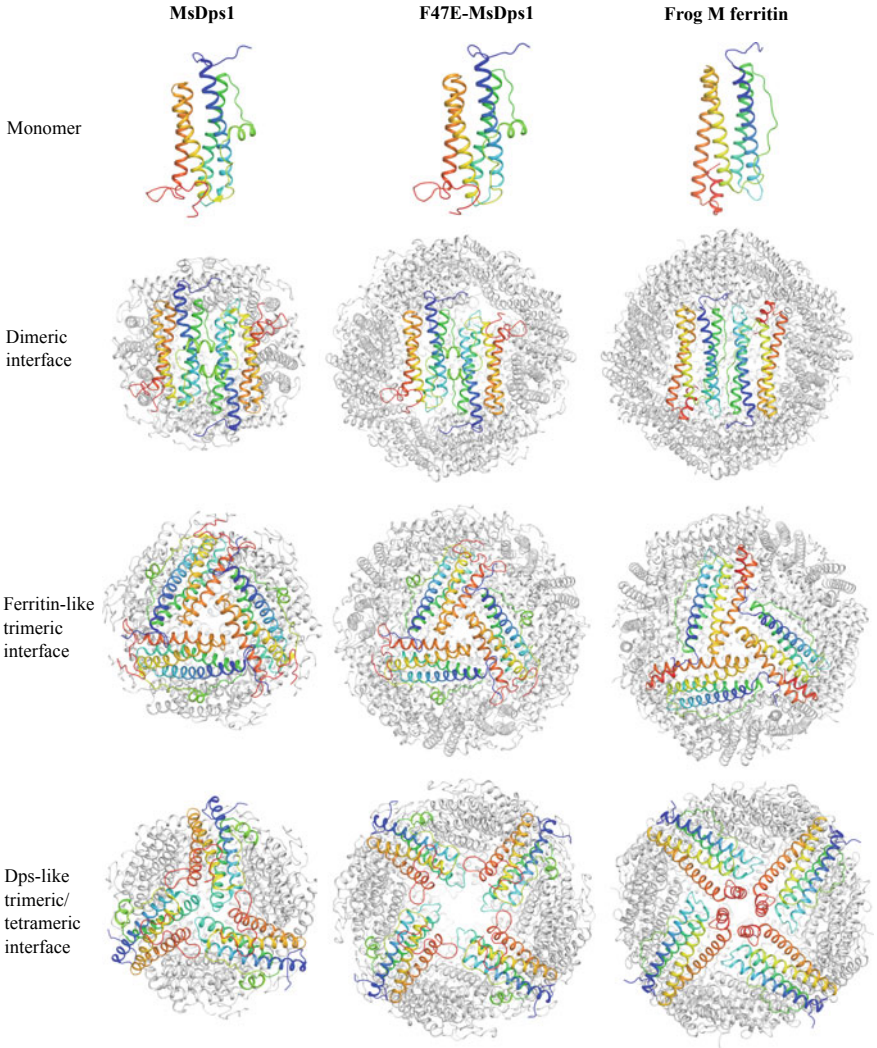


Fig. 3.20 *Symmetry related interfaces in MsDps1 (PDB ID: 1VEI), F47E-MsDps1 (PDB ID: 5H46) and frog M ferritin (PDB ID: 3RBC).* F47E-MsDps1 has an arrangement similar to ferritins. Dps proteins have a dps-like trimeric interface, distinct from ferritins and F47E-MsDps1which has a tetrameric interface. In Ferritins, E helix is at the centre of the tetrameric interface. F47E-MsDps1 has a large gap in the assembly of the tetramer having no E helices

conserved phenyl alanine to a glutamate (F47E) in the AB loop of *M. smegmatis* Dps1 was shown to alter its assembly from the canonical 12-mer to a ferritin-like 24-mer. The assembly of mutated Dps monomers to ferritin-like cages seems to be stabilised only under crystallization conditions, as in solution these subunits exist as dimers (Williams et al. 2017).

The crystal structure of F47E-MsDps1 exhibited a 24-meric ferritin-like assembly, with an outer diameter of 120 Å and an inner diameter of 80 Å similar to ferritins. WT Dps has an outer diameter of 90 Å and a central cavity of 45 Å. The F47E-MsDps1 had a 432 octahedral symmetry, with four-fold, ferritin-like three-fold and two-fold symmetries (Fig. 3.20). The point mutation F47E caused the AB loop to lose its rigid conformation due to disruption of the hydrophobic interactions at the dps-like three-fold, pushing its conversion to a four-fold interface similar to ferritins (Fig. 3.21). In ferritins, a C-terminal E helix is poised at the centre of the four-fold and stabilises the tetrameric interface. It is interesting to note that some ferritin homologs have a short AB loop and others have an AB loop which faces away from the four-fold interface, so as to avoid clashes with the E-helix.

Ferritin family proteins are thought to have originated from rubyerythrin-like ancestors which evolved into 12-meric bacterioferritins (Bfr) that later diverged into 12-meric Dps and 24-meric ferritins/Bfrs (Andrews 2010). Thus, early ferritin-like proteins could have had a dodecameric assembly, and as indicated by the mutational switch F47E, minor changes in secondary structure could have led to the evolution of 24-mers.

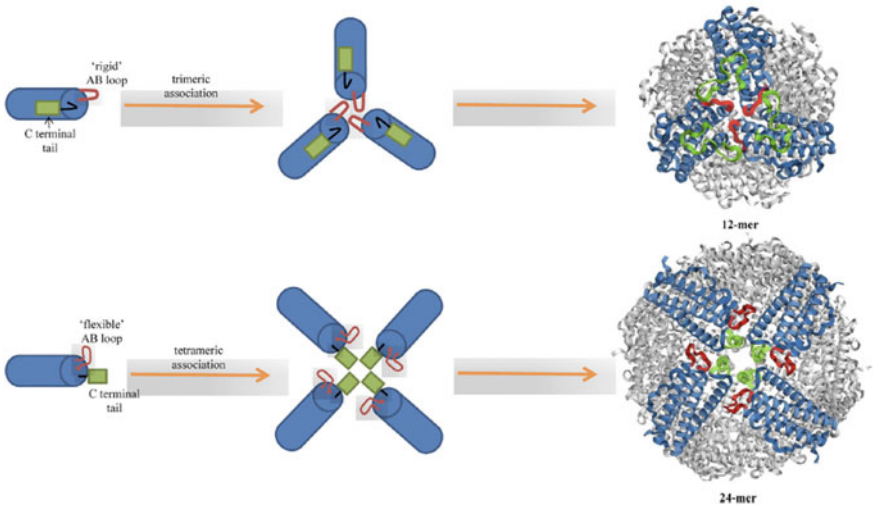


Fig. 3.21 *Model for assembly of ferritin family proteins.* Top panel shows the proposed assembly of a Dps 12-mer, where a four-helix monomer with a "rigid" AB loop associates via the dps-like trimeric interface. Bottom panel shows the monomers associating through a tetrameric interface with the C-terminal E helix at the centre, forming 24-meric ferritins. Figure reproduced with permission from Structure. 2017. 25: 1449–1454

Applications in Biotechnology and Nanomedicine

Biomolecules like Dps proteins with cage-like assemblies can be engineered via its inner shell, external surface or subunit interfaces. The hollow interior can be manipulated to load therapeutic drugs and other diagnostic molecules. External surface modifications can help to enhance biocompatibility and cell-specific targeting. Self-assembly can be modulated through altering the subunit interfaces. The iron release pores can be altered to fine tune release of molecular cargos (He and Marles-Wright 2015). This section further explores the applications of Dps in biotechnology, drug delivery and nonomedicine.

Engineering of Nanoscale Devices

Multiprotein complexes like Dps with their well-defined interior cavities are attractive templates for bionanotechnology. They can be produced in large quantities, are amenable to chemical modifications, are structurally and functionally well characterised (He and Marles-Wright 2015). The synthetic flexibility of these miniferritins have been demonstrated by computational design introducing wide-scale mutations for altering polarity of the interior from a hydrophilic to hydrophobic cavity. These mutated Dps could assemble similar to WT proteins and form dodecamers, paving way for engineering proteins with novel functions (Swift et al. 2006). Dps can bind other metal ions in addition to iron and this property has been utilized to generate mineral cores of cobalt and oxygen in *L. innocua* Dps (Allen et al. 2003).

Ferritin-like proteins are used to mineralise non-physiological metals through self-assembly around a solution of metal ions, chemically mediated redox reactions or photochemistry (Yoshimura 2006). Ferritins can also mineralise iron sulphide, Mn(III) and magnetite (Fe_3O_4) cores. Ferritin-like proteins with magnetite core called magnetoferritins have uses in cell imaging, providing magnetic contrast and magnetic separation of cells and particles (Deans et al. 2006).

Dps has been used to fabricate inorganic nanodevices like quantum dots and nanowires. Here, Dps cores loaded with a mineral core are deposited on silicon substrates which are silanized and made functional with small peptides or molecules. The protein cage can be removed with heat to leave the core in place, and such semiconductor cores deposited on silicon wafers can be used as memory gates (Okuda et al. 2005). Metals cores isolated from Dps can be used to seed the growth of carbon non-tubes and nanowires (Kim et al. 2011). Dps loaded with mineral nanoparticles can crystallise in 2-dimensional arrays to give a regular array of uniformly sized nanoparticles on solid surfaces.

Due to the capacity for mineralisation of heavy atoms in the core, Dps can be used as contrast agents in electron microscopy. The stability of Dps and its ordered symmetric structure makes it an ideal sample for cryoEM imaging techniques. Apoferritins are routinely used as a standard for estimation of the magnification of micrographs in cryoEM (Wasilewski et al. 2012).

Nanoparticles in Biomedicine

Ferritin family proteins are stable and biocompatible entities that can be manipulated with respect to their assembly, surface modifications and reconstitution. Thus, they have the potential to be developed as drug delivery vehicles. Due to their hydrophilic cavities, incorporation of non-metal containing drugs poses challenges. Addition of charged molecules or drugs complexed with metals such as copper, could overcome this limitation (Maham et al. 2009, 2011). Surface modifications with epitopes and labels to target these loaded nanovehicles to specific cell types or tumours, could be a powerful tool for drug delivery.

Ferritins have been used to develop self-assembling nanoparticles that elicit broader and more potent immunity in comparison to traditional vaccines. In this instance, haemagglutinin from the influenza virus was fused to ferritins, which assembled via the trimeric ferritin-like interfaces to give eight trimeric viral spikes on the ferritin surface. Immunization with the HA-ferritin fusion protein generated an increased HA inhibition antibody titre compared to the traditional inactivated vaccines (Kanekiyo et al. 2013). Also, there was no autoimmune reaction here, due to the use of *H. pylori* ferritin with high sequence variation from the human ferritin. Similar fusion proteins have been engineered with Dps, with a roles in diagnosis, treatment and prevention of diseases. Dps fusion proteins are soluble, in spite of the poor solubility of the attached peptide/epitope. They also have enhanced thermostability and enhanced immune response due to multimeric assembly, similar to ferritins (Verma et al. 2018).

Biomoleules can often rival synthetic molecules due to their flexibility, biocompatibility, diversity and lower toxicity. Naturally occurring nanocompartments like Dps can thus be exploited to encapsulate nanoparticles, and surface modifications can help to target tissue systems and make them suitable as drug carriers. Understanding the loading and unloading mechanisms, as exemplified in the study of gating mechanisms in Dps can help us understand how to fine tune the release of drugs in a tissue-specific manner (Williams et al. 2014). Also, having a better knowledge of multi-subunit assembly of the ferritin family proteins can help in building higher order assemblies (Williams et al. 2017).

References

Alaleona F, Franceschini S, Ceci P, Ilari A, Chiancone E (2010) Thermosynechococcus elongatus DpsA binds Zn(II) at a unique three histidine-containing ferroxidase center and utilizes O_2 as iron oxidant with very high efficiency, unlike the typical Dps proteins. FEBS J 277(4):903–917. https://doi.org/10.1111/j.1742-4658.2009.07532.x

Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J Bacteriol 181(20):6361–6370

Allen M, Willits D, Young M, Douglas T (2003) Constrained synthesis of cobalt oxide nanomaterials in the 12-subunit protein cage from Listeria innocua. Inorg Chem 42(20):6300–6305. https://doi.org/10.1021/ic0343657

Almiron M, Link AJ, Furlong D, Kolter R (1992) A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. Genes Dev 6(12B):2646–2654. https://doi.org/10.1101/gad.6.12b.2646

Altuvia S, Almiron M, Huisman G, Kolter R, Storz G (1994) The dps promoter is activated by OxyR during growth and by IHF and sigma S in stationary phase. Mol Microbiol 13(2):265–272. https://doi.org/10.1111/j.1365-2958.1994.tb00421.x

Andrews SC (2010) The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. Biochim Biophys Acta 1800(8):691–705. https://doi.org/10.1016/j.bbagen.2010.05.010

Ardini M, Fiorillo A, Fittipaldi M, Stefanini S, Gatteschi D, Ilari A, Chiancone E (2013) Kineococcus radiotolerans Dps forms a heteronuclear Mn-Fe ferroxidase center that may explain the Mn-dependent protection against oxidative stress. Biochim Biophys Acta 1830(6):3745–3755. https://doi.org/10.1016/j.bbagen.2013.02.003

Arnold AR, Barton JK (2013) DNA protection by the bacterial ferritin Dps via DNA charge transport. J Am Chem Soc 135(42):15726–15729. https://doi.org/10.1021/ja408760w

Bellapadrona G, Ardini M, Ceci P, Stefanini S, Chiancone E (2010) Dps proteins prevent Fenton-mediated oxidative damage by trapping hydroxyl radicals within the protein shell. Free Radic Biol Med 48(2):292–297. https://doi.org/10.1016/j.freeradbiomed.2009.10.053

Bellapadrona G, Stefanini S, Zamparelli C, Theil EC, Chiancone E (2009) Iron translocation into and out of Listeria innocua Dps and size distribution of the protein-enclosed nanomineral are modulated by the electrostatic gradient at the 3-fold "ferritin-like" pores. J Biol Chem 284(28):19101–19109. https://doi.org/10.1074/jbc.M109.014670

Bhattacharyya G, Grove A (2007) The N-terminal extensions of Deinococcus radiodurans Dps-1 mediate DNA major groove interactions as well as assembly of the dodecamer. J Biol Chem 282(16):11921–11930. https://doi.org/10.1074/jbc.M611255200

Bozzi M, Mignogna G, Stefanini S, Barra D, Longhi C, Valenti P, Chiancone E (1997) A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in Listeria innocua. J Biol Chem 272(6):3259–3265. https://doi.org/10.1074/jbc.272.6.3259

Calhoun LN, Kwon YM (2011) Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. J Appl Microbiol 110(2):375–386. https://doi.org/10.1111/j.1365-2672.2010.04890.x

Castruita M, Saito M, Schottel PC, Elmegreen LA, Myneni S, Stiefel EI, Morel FM (2006) Overexpression and characterization of an iron storage and DNA-binding Dps protein from Trichodesmium erythraeum. Appl Environ Microbiol 72(4):2918–2924. https://doi.org/10.1128/AEM.72.4.2918-2924.2006

Ceci P, Cellai S, Falvo E, Rivetti C, Rossi GL, Chiancone E (2004) DNA condensation and self-aggregation of Escherichia coli Dps are coupled phenomena related to the properties of the N-terminus. Nucleic Acids Res 32(19):5935–5944. https://doi.org/10.1093/nar/gkh915

Ceci P, Ilari A, Falvo E, Chiancone E (2003) The Dps protein of Agrobacterium tumefaciens does not bind to DNA but protects it toward oxidative cleavage: x-ray crystal structure, iron binding, and hydroxyl-radical scavenging properties. J Biol Chem 278(22):20319–20326. https://doi.org/10.1074/jbc.M302114200

Ceci P, Mangiarotti L, Rivetti C, Chiancone E (2007) The neutrophil-activating Dps protein of Helicobacter pylori, HP-NAP, adopts a mechanism different from Escherichia coli Dps to bind and condense DNA. Nucleic Acids Res 35(7):2247–2256. https://doi.org/10.1093/nar/gkm077

Chiancone E, Ceci P (2010) Role of Dps (DNA-binding proteins from starved cells) aggregation on DNA. Front Biosci (Landmark Ed) 15:122–131. https://doi.org/10.2741/3610

Chowdhury RP, Gupta S, Chatterji D (2007) Identification and characterization of the dps promoter of Mycobacterium smegmatis: promoter recognition by stress-specific extracytoplasmic function sigma factors sigmaH and sigmaF. J Bacteriol 189(24):8973–8981. https://doi.org/10.1128/JB.01222-07

Deans AE, Wadghiri YZ, Bernas LM, Yu X, Rutt BK, Turnbull DH (2006) Cellular MRI contrast via coexpression of transferrin receptor and ferritin. Magn Reson Med 56(1):51–59. https://doi.org/10.1002/mrm.20914

Durham KA, Bullerjahn GS (2002) Immunocytochemical localization of the stress-induced DpsA protein in the cyanobacterium Synechococcus sp. strain PCC 7942. J Basic Microbiol 42 (6):367–372. doi:https://doi.org/10.1002/1521-4028(200212)42:6<367::AID-JOBM367>3.0.CO;2-T

Facey PD, Hitchings MD, Saavedra-Garcia P, Fernandez-Martinez L, Dyson PJ, Del Sol R (2009) Streptomyces coelicolor Dps-like proteins: differential dual roles in response to stress during vegetative growth and in nucleoid condensation during reproductive cell division. Mol Microbiol 73(6):1186–1202. https://doi.org/10.1111/j.1365-2958.2009.06848.x

Fenton HJH (1894) LXXIII-Oxidation of tartaric acid in presence of iron. J Chem Soc Trans 65:899–910. https://doi.org/10.1039/ct8946500899

Franceschini S, Ceci P, Alaleona F, Chiancone E, Ilari A (2006) Antioxidant Dps protein from the thermophilic cyanobacterium Thermosynechococcus elongatus. FEBS J 273(21):4913–4928. https://doi.org/10.1111/j.1742-4658.2006.05490.x

Frenkiel-Krispin D, Levin-Zaidman S, Shimoni E, Wolf SG, Wachtel EJ, Arad T, Finkel SE, Kolter R, Minsky A (2001) Regulated phase transitions of bacterial chromatin: a non-enzymatic pathway for generic DNA protection. EMBO J 20(5):1184–1191. https://doi.org/10.1093/emboj/20.5.1184

Ganguly A, Rajdev P, Williams SM, Chatterji D (2012) Nonspecific interaction between DNA and protein allows for cooperativity: a case study with mycobacterium DNA binding protein. J Phys Chem B 116(1):621–632. https://doi.org/10.1021/jp209423n

Gao JL, Lu Y, Browne G, Yap BC, Trewhella J, Hunter N, Nguyen KA (2012) The role of heme binding by DNA-protective protein from starved cells (Dps) in the Tolerance of Porphyromonas gingivalis to heme toxicity. J Biol Chem 287(50):42243–42258. https://doi.org/10.1074/jbc.M112.392787

Gaupp R, Ledala N, Somerville GA (2012) Staphylococcal response to oxidative stress. Front Cell Infect Microbiol 2:33. https://doi.org/10.3389/fcimb.2012.00033

Ghatak P, Karmakar K, Kasetty S, Chatterji D (2011) Unveiling the role of Dps in the organization of mycobacterial nucleoid. PLoS ONE 6(1):e16019. https://doi.org/10.1371/journal.pone.0016019

Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nat Struct Biol 5(4):294–303. https://doi.org/10.1038/nsb0498-294

Gupta S, Chatterji D (2003) Bimodal protection of DNA by Mycobacterium smegmatis DNA-binding protein from stationary phase cells. J Biol Chem 278(7):5235–5241. https://doi.org/10.1074/jbc.M208825200

Haikarainen T, Thanassoulas A, Stavros P, Nounesis G, Haataja S, Papageorgiou AC (2011) Structural and thermodynamic characterization of metal ion binding in Streptococcus suis Dpr. J Mol Biol 405(2):448–460. https://doi.org/10.1016/j.jmb.2010.10.058

Haikarainen T, Tsou CC, Wu JJ, Papageorgiou AC (2010) Structural characterization and biological implications of di-zinc binding in the ferroxidase center of Streptococcus pyogenes Dpr. Biochem Biophys Res Commun 398(3):361–365. https://doi.org/10.1016/j.bbrc.2010.06.071

Havukainen H, Haataja S, Kauko A, Pulliainen AT, Salminen A, Haikarainen T, Finne J, Papageorgiou AC (2008) Structural basis of the zinc- and terbium-mediated inhibition of ferroxidase activity in Dps ferritin-like proteins. Protein Sci 17(9):1513–1521. https://doi.org/10.1110/ps.036236.108

Hayden JA, Hendrich MP (2010) EPR spectroscopy and catalase activity of manganese-bound DNA-binding protein from nutrient starved cells. J Biol Inorg Chem 15(5):729–736. https://doi.org/10.1007/s00775-010-0640-3

He D, Marles-Wright J (2015) Ferritin family proteins and their use in bionanotechnology. N Biotechnol 32(6):651–657. https://doi.org/10.1016/j.nbt.2014.12.006

Hong Y, Wang G, Maier RJ (2006) Helicobacter hepaticus Dps protein plays an important role in protecting DNA from oxidative damage. Free Radic Res 40(6):597–605. https://doi.org/10.1080/10715760600618882

Howe C, Ho F, Nenninger A, Raleiras P, Stensjo K (2018) Differential biochemical properties of three canonical Dps proteins from the cyanobacterium Nostoc punctiforme suggest distinct cellular functions. J Biol Chem 293(43):16635–16646. https://doi.org/10.1074/jbc.RA118.002425

Ilari A, Stefanini S, Chiancone E, Tsernoglou D (2000) The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. Nat Struct Biol 7(1):38–43. https://doi.org/10.1038/71236

Ishikawa T, Mizunoe Y, Kawabata S, Takade A, Harada M, Wai SN, Yoshida S (2003) The iron-binding protein Dps confers hydrogen peroxide stress resistance to Campylobacter jejuni. J Bacteriol 185(3):1010–1017. https://doi.org/10.1128/jb.185.3.1010-1017.2003

Janissen R, Arens MMA, Vtyurina NN, Rivai Z, Sunday ND, Eslami-Mossallam B, Gritsenko AA, Laan L, de Ridder D, Artsimovitch I, Dekker NH, Abbondanzieri EA, Meyer AS (2018) Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. Cell 174 (5):1188–1199 e1114. doi:https://doi.org/10.1016/j.cell.2018.06.049

Jeong KC, Hung KF, Baumler DJ, Byrd JJ, Kaspar CW (2008) Acid stress damage of DNA is prevented by Dps binding in Escherichia coli O157:H7. BMC Microbiol 8:181. https://doi.org/10.1186/1471-2180-8-181

Kamyshinsky R, Chesnokov Y, Dadinova L, Mozhaev A, Orlov I, Petoukhov M, Orekhov A, Shtykova E, Vasiliev A (2019) Polymorphic Protective Dps-DNA Co-Crystals by Cryo Electron Tomography and Small Angle X-Ray Scattering. Biomolecules 10 (1). doi:https://doi.org/10.3390/biom10010039

Kanekiyo M, Wei CJ, Yassine HM, McTamney PM, Boyington JC, Whittle JR, Rao SS, Kong WP, Wang L, Nabel GJ (2013) Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. Nature 499(7456):102–106. https://doi.org/10.1038/nature12202

Kauko A, Pulliainen AT, Haataja S, Meyer-Klaucke W, Finne J, Papageorgiou AC (2006) Iron incorporation in Streptococcus suis Dps-like peroxide resistance protein Dpr requires mobility in the ferroxidase center and leads to the formation of a ferrihydrite-like core. J Mol Biol 364(1):97–109. https://doi.org/10.1016/j.jmb.2006.08.061

Kim H-J, Oh E, Lee J et al (2011) Synthesis of carbon nanotubes with catalytic iron-containing proteins. Carbon 49(12):3717–3722. https://doi.org/10.1016/j.carbon.2011.04.037

Kottakis F, Papadopoulos G, Pappa EV, Cordopatis P, Pentas S, Choli-Papadopoulou T (2008) Helicobacter pylori neutrophil-activating protein activates neutrophils by its C-terminal region even without dodecamer formation, which is a prerequisite for DNA protection–novel approaches against Helicobacter pylori inflammation. FEBS J 275(2):302–317. https://doi.org/10.1111/j.1742-4658.2007.06201.x

Li X, Pal U, Ramamoorthi N, Liu X, Desrosiers DC, Eggers CH, Anderson JF, Radolf JD, Fikrig E (2007) The Lyme disease agent Borrelia burgdorferi requires BB0690, a Dps homologue, to persist within ticks. Mol Microbiol 63(3):694–710. https://doi.org/10.1111/j.1365-2958.2006.05550.x

Linder MC (2013) Mobilization of stored iron in mammals: a review. Nutrients 5(10):4022–4050. https://doi.org/10.3390/nu5104022

Luijsterburg MS, Noom MC, Wuite GJ, Dame RT (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. J Struct Biol 156(2):262–272. https://doi.org/10.1016/j.jsb.2006.05.006

Maham A, Tang Z, Wu H, Wang J, Lin Y (2009) Protein-based nanomedicine platforms for drug delivery. Small 5(15):1706–1721. https://doi.org/10.1002/smll.200801602

Maham A, Wu H, Wang J, Kang X, Zhang Y, Lin Y (2011) Apoferritin-based nanomedicine platform for drug delivery: equilibrium binding study of daunomycin with DNA. J Mater Chem 21:8700. https://doi.org/10.1039/c0jm04321d

Martinez A, Kolter R (1997) Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. J Bacteriol 179(16):5188–5194. https://doi.org/10.1128/jb.179.16.5188-5194.1997

Morikawa K, Ohniwa RL, Kim J, Maruyama A, Ohta T, Takeyasu K (2006) Bacterial nucleoid dynamics: oxidative stress response in Staphylococcus aureus. Genes Cells 11(4):409–423. https://doi.org/10.1111/j.1365-2443.2006.00949.x

Nair S, Finkel SE (2004) Dps protects cells against multiple stresses during stationary phase. J Bacteriol 186(13):4192–4198. https://doi.org/10.1128/JB.186.13.4192-4198.2004

Narayan OP, Kumari N, Rai LC (2010) Heterologous expression of Anabaena PCC 7120 all3940 (a Dps family gene) protects Escherichia coli from nutrient limitation and abiotic stresses. Biochem Biophys Res Commun 394(1):163–169. https://doi.org/10.1016/j.bbrc.2010.02.135

Nicodeme M, Perrin C, Hols P, Bracquart P, Gaillard JL (2004) Identification of an iron-binding protein of the Dps family expressed by Streptococcus thermophilus. Curr Microbiol 48(1):51–56. https://doi.org/10.1007/s00284-003-4116-3

Okuda M, Kobayashi Y, Suzuki K, Sonoda K, Kondoh T, Wagawa A, Kondo A, Yoshimura H (2005) Self-organized inorganic nanoparticle arrays on protein lattices. Nano Lett 5(5):991–993. https://doi.org/10.1021/nl050556q

Pacello F, Ceci P, Ammendola S, Pasquali P, Chiancone E, Battistoni A (2008) Periplasmic Cu, Zn superoxide dismutase and cytoplasmic Dps concur in protecting Salmonella enterica serovar Typhimurium from extracellular reactive oxygen species. Biochim Biophys Acta 1780(2):226–232. https://doi.org/10.1016/j.bbagen.2007.12.001

Papinutto E, Dundon WG, Pitulis N, Battistutta R, Montecucco C, Zanotti G (2002) Structure of two iron-binding proteins from Bacillus anthracis. J Biol Chem 277(17):15093–15098. https://doi.org/10.1074/jbc.M112378200

Park M, Yun ST, Hwang SY, Chun CI, Ahn TI (2006) The dps gene of symbiotic "Candidatus Legionella jeonii" in Amoeba proteus responds to hydrogen peroxide and phagocytosis. J Bacteriol 188(21):7572–7580. https://doi.org/10.1128/JB.00576-06

Pena MM, Bullerjahn GS (1995) The DpsA protein of Synechococcus sp. Strain PCC7942 is a DNA-binding hemoprotein. Linkage of the Dps and bacterioferritin protein families. J Biol Chem 270 (38):22478–22482. doi:https://doi.org/10.1074/jbc.270.38.22478

Pesek J, Buchler R, Albrecht R, Boland W, Zeth K (2011) Structure and mechanism of iron translocation by a Dps protein from Microbacterium arborescens. J Biol Chem 286(40):34872–34882. https://doi.org/10.1074/jbc.M111.246108

Pulliainen AT, Kauko A, Haataja S, Papageorgiou AC, Finne J (2005) Dps/Dpr ferritin-like protein: insights into the mechanism of iron incorporation and evidence for a central role in cellular iron homeostasis in Streptococcus suis. Mol Microbiol 57(4):1086–1100. https://doi.org/10.1111/j.1365-2958.2005.04756.x

Ren B, Tibbelin G, Kajino T, Asami O, Ladenstein R (2003) The multi-layered structure of Dps with a novel di-nuclear ferroxidase center. J Mol Biol 329(3):467–477. https://doi.org/10.1016/s0022-2836(03)00466-2

Reon BJ, Nguyen KH, Bhattacharyya G, Grove A (2012) Functional comparison of Deinococcus radiodurans Dps proteins suggests distinct in vivo roles. Biochem J 447(3):381–391. https://doi.org/10.1042/BJ20120902

Rocha ER, Herren CD, Smalley DJ, Smith CJ (2003) The complex oxidative stress response of Bacteroides fragilis: the role of OxyR in control of gene expression. Anaerobe 9(4):165–173. https://doi.org/10.1016/S1075-9964(03)00118-5

Roy S, Saraswathi R, Gupta S, Sekar K, Chatterji D, Vijayan M (2007) Role of N and C-terminal tails in DNA binding and assembly in Dps: structural studies of Mycobacterium smegmatis Dps deletion mutants. J Mol Biol 370(4):752–767. https://doi.org/10.1016/j.jmb.2007.05.004

Santos SP, Mitchell EP, Franquelim HG, Castanho MA, Abreu IA, Romao CV (2015) Dps from Deinococcus radiodurans: oligomeric forms of Dps1 with distinct cellular functions and Dps2 involved in metal storage. FEBS J 282(22):4307–4327. https://doi.org/10.1111/febs.13420

Schmidt R, Zahn R, Bukau B, Mogk A (2009) ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. Mol Microbiol 72(2):506–517. https://doi.org/10.1111/j.1365-2958.2009.06666.x

Schwartz JK, Liu XS, Tosha T, Diebold A, Theil EC, Solomon EI (2010) CD and MCD spectroscopic studies of the two Dps miniferritin proteins from Bacillus anthracis: role of O2 and H2O2 substrates in reactivity of the diiron catalytic centers. Biochemistry 49(49):10516–10525. https://doi.org/10.1021/bi101346c

Shu JC, Soo PC, Chen JC, Hsu SH, Chen LC, Chen CY, Liang SH, Buu LM, Chen CC (2013) Differential regulation and activity against oxidative stress of Dps proteins in Bacillus cereus. Int J Med Microbiol 303(8):662–673. https://doi.org/10.1016/j.ijmm.2013.09.011

Stefanini S, Cavallo S, Montagnini B, Chiancone E (1999) Incorporation of iron by the unusual dodecameric ferritin from Listeria innocua. Biochem J 338(Pt 1):71–75

Stephani K, Weichart D, Hengge R (2003) Dynamic control of Dps protein levels by ClpXP and ClpAP proteases in Escherichia coli. Mol Microbiol 49(6):1605–1614. https://doi.org/10.1046/j.1365-2958.2003.03644.x

Stillman TJ, Upadhyay M, Norte VA, Sedelnikova SE, Carradus M, Tzokov S, Bullough PA, Shearman CA, Gasson MJ, Williams CH, Artymiuk PJ, Green J (2005) The crystal structures of Lactococcus lactis MG1363 Dps proteins reveal the presence of an N-terminal helix that is required for DNA binding. Mol Microbiol 57(4):1101–1112. https://doi.org/10.1111/j.1365-2958.2005.04757.x

Su M, Cavallo S, Stefanini S, Chiancone E, Chasteen ND (2005) The so-called Listeria innocua ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties. Biochemistry 44 (15):5572–5578. doi:https://doi.org/10.1021/bi0472705

Swift J, Wehbi WA, Kelly BD, Stowell XF, Saven JG, Dmochowski IJ (2006) Design of functional ferritin-like proteins with hydrophobic cavities. J Am Chem Soc 128(20):6611–6619. https://doi.org/10.1021/ja057069x

Theoret JR, Cooper KK, Zekarias B, Roland KL, Law BF, Curtiss R 3rd, Joens LA (2012) The Campylobacter jejuni Dps homologue is important for in vitro biofilm formation and cecal colonization of poultry and may serve as a protective antigen for vaccination. Clin Vaccine Immunol 19(9):1426–1431. https://doi.org/10.1128/CVI.00151-12

Travers A, Muskhelishvili G (2005) Bacterial chromatin. Curr Opin Genet Dev 15(5):507–514. https://doi.org/10.1016/j.gde.2005.08.006

Tsou CC, Chiang-Ni C, Lin YS, Chuang WJ, Lin MT, Liu CC, Wu JJ (2008) An iron-binding protein, Dpr, decreases hydrogen peroxide stress and protects Streptococcus pyogenes against multiple stresses. Infect Immun 76(9):4038–4045. https://doi.org/10.1128/IAI.00477-08

Verma D, Gulati N, Kaul S, Mukherjee S, Nagaich U (2018) Protein Based Nanostructures for Drug Delivery. J Pharm (Cairo) 2018:9285854. https://doi.org/10.1155/2018/9285854

Vtyurina NN, Dulin D, Docter MW, Meyer AS, Dekker NH, Abbondanzieri EA (2016) Hysteresis in DNA compaction by Dps is described by an Ising model. Proc Natl Acad Sci U S a 113(18):4982–4987. https://doi.org/10.1073/pnas.1521241113

Wasilewski S, Karelina D, Berriman JA, Rosenthal PB (2012) Automatic magnification determination of electron cryomicroscopy images using apoferritin as a standard. J Struct Biol 180(1):243–248. https://doi.org/10.1016/j.jsb.2012.07.006

Williams SM, Chatterji D (2017) Flexible aspartates propel iron to the ferroxidation sites along pathways stabilized by a conserved arginine in Dps proteins from Mycobacterium smegmatis. Metallomics 9(6):685–698. https://doi.org/10.1039/c7mt00008a

Williams SM, Chandran AV, Prakash S, Vijayan M, Chatterji D (2017) A Mutation Directs the Structural Switch of DNA Binding Proteins under Starvation to a Ferritin-like Protein Cage. Structure 25 (9):1449–1454 e1443. doi:https://doi.org/10.1016/j.str.2017.07.006

Williams SM, Chandran AV, Vijayabaskar MS, Roy S, Balaram H, Vishveshwara S, Vijayan M, Chatterji D (2014) A histidine aspartate ionic lock gates the iron passage in miniferritins from Mycobacterium smegmatis. J Biol Chem 289(16):11042–11058. https://doi.org/10.1074/jbc.M113.524421

Wolf SG, Frenkiel D, Arad T, Finkel SE, Kolter R, Minsky A (1999) DNA protection by stress-induced biocrystallization. Nature 400(6739):83–85. https://doi.org/10.1038/21918

Yamamoto Y, Poole LB, Hantgan RR, Kamio Y (2002) An iron-binding protein, Dpr, from Streptococcus mutans prevents iron-dependent hydroxyl radical formation in vitro. J Bacteriol 184(11):2931–2939. https://doi.org/10.1128/jb.184.11.2931-2939.2002

Yokoyama H, Fujii S (2014) Structures and metal-binding properties of Helicobacter pylori neutrophil-activating protein with a di-nuclear ferroxidase center. Biomolecules 4(3):600–615. https://doi.org/10.3390/biom4030600

Yoshimura H (2006) Protein-assisted nanoparticle synthesis. Colloids Surf a Physicochem Eng Asp 282–283:464–470. https://doi.org/10.1016/j.colsurfa.2006.01.037

Yu MJ, Ren J, Zeng YL, Zhou SN, Lu YJ (2009) The Legionella pneumophila Dps homolog is regulated by iron and involved in multiple stress tolerance. J Basic Microbiol 49(Suppl 1):S79-86. https://doi.org/10.1002/jobm.200800357

Zeth K, Offermann S, Essen LO, Oesterhelt D (2004) Iron-oxo clusters biomineralizing on protein surfaces: structural analysis of Halobacterium salinarum DpsA in its low- and high-iron states. Proc Natl Acad Sci U S a 101(38):13780–13785. https://doi.org/10.1073/pnas.0401821101

Zhao G, Ceci P, Ilari A, Giangiacomo L, Laue TM, Chiancone E, Chasteen ND (2002) Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of Escherichia coli. J Biol Chem 277 (31):27689–27696. doi:https://doi.org/10.1074/jbc.M202094200

Chapter 4
Structure and Function of the Separase-Securin Complex

Shukun Luo and Liang Tong

Abstract Separase is a large cysteine protease in eukaryotes and has crucial roles in many cellular processes, especially chromosome segregation during mitosis and meiosis, apoptosis, DNA damage repair, centrosome disengagement and duplication, spindle stabilization and elongation. It dissolves the cohesion between sister chromatids by cleaving one of the subunits of the cohesin ring for chromosome segregation. The activity of separase is tightly controlled at many levels, through direct binding of inhibitory proteins as well as posttranslational modification. Dysregulation of separase activity is linked to cancer and genome instability, making it a target for drug discovery. One of the best-known inhibitors of separase is securin, which has been identified in yeast, plants, and animals. Securin forms a tight complex with separase and potently inhibits its catalytic activity. Recent structures of the separase-securin complex have revealed the molecular mechanism for the inhibitory activity of securin. A segment of securin is bound in the active site of separase, thereby blocking substrate binding. Securin itself is not cleaved by separase as its binding mode is not compatible with catalysis. Securin also has extensive interactions with separase outside the active site, consistent with its function as a chaperone to stabilize this enzyme.

Keywords Sister chromatids · Cohesin · Mitosis · Meiosis · DNA damage repair · Cancer

S. Luo · L. Tong (✉)
Department of Biological Sciences, Columbia University, New York, NY 10027, USA
e-mail: ltong@columbia.edu

S. Luo
State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: shukunluo@sjtu.edu.cn

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_4

Introduction

Precise segregation of chromosomes is a prerequisite of faithful cell division and hence it is a tightly regulated process. During DNA synthesis in the interphase, sister chromatids are paired to each other through cohesion by the ring-shaped cohesin complex. Discovered in the late 1990s, the cohesin ring is primarily composed of three subunits: Smc1, Smc3 and a bridging subunit Scc1 between them, and it is also involved in DNA loop extrusion (Davidson et al. 2019; Kim et al. 2019). Two mechanisms are mainly responsible for removing cohesins from chromosomes during cell division. The first is through the prophase pathway, by which the majority of cohesins located on chromosome arms are removed during prophase, a process involving the Wapl protein, polo-like kinase Plk1, and Aurora B kinase (Sumara et al. 2002; Kueng et al. 2006; Gimenez-Abian et al. 2004; Hauf et al. 2005; Gandhi et al. 2006). However, cohesins close to the centromere region are protected from the prophase pathway and are removed instead during the transition from metaphase to anaphase by the protease separase, thus this pathway is known as the separase pathway. Separase cleaves the Scc1 subunit at two specific sites and opens the cohesin ring to resolve chromosome cohesion in a spatial and timely manner.

Separase was first discovered by yeast genetics in the late 1980s and early 1990s, and its loss of function was related to spindle pole body and polyploid nuclei abnormalities (Uzawa et al. 1990; Baum et al. 1988; May et al. 1992). Separase was further revealed to be the protease that targets the cohesin subunit Scc1/Mcd1 for chromosome resolution (Buonomo et al. 2000; Uhlmann et al. 2000; Waizenegger et al. 2000). Multiple other functional roles have been discovered since then, including apoptosis regulation (Hellmuth and Stemmann 2020), DNA damage repair (Nagao et al. 2004; McAleenan et al. 2013; Hellmuth et al. 2018) and centrosome disengagement and duplication (Tsou et al. 2009; Schockel et al. 2011; Lee and Rhee 2012; Matsuo et al. 2012; Nakamura et al. 2009). Due to its crucial roles in important cellular processes, separase is tightly regulated at many levels, including the inhibition by two well-known inhibitors securin and CDK1-cyclin B1 (Ciosk et al. 1998; Zou et al. 1999; Stemmann et al. 2001; Gorr et al. 2005; Karasu et al. 2019; Li et al. 2019; Holland et al. 2007) and a recently discovered inhibitor shugoshin-MAD2 complex (Hellmuth et al. 2020). Dysregulation of separase leads to aneuploidy and chromosome instability, which has been linked to diseases including cancers (Zhang et al. 2008; Mukherjee et al. 2014a). The molecular mechanisms for the activity and regulation of separase have been unraveled recently by several structural studies (Viadiu et al. 2005; Luo and Tong 2017; Lin et al. 2016; Boland et al. 2017). In this chapter, we discuss the molecular characterization, functional roles, regulation and the structure of separase and its complex with securin.

Domain Organization of the Cysteine Protease Separase

Separase is an endopeptidase that belongs to the CD clan of cysteine proteases, related to the well-known caspases in the same clan. It's substrate contains an arginine at the P_1 position before the scissile bond and several additional residues are conserved for recognition (Lin et al. 2016). The size of separase varies substantially, ranging from 140 to 250 kD in eukaryotes. Sequence analysis shows that the C-terminal region where the catalytic domain is located (Fig. 4.1) has the highest conservation across species. Sequence diversity is observed in the N-terminal region and the unstructured segment in the middle, which contributes to the size variations among separases. Atypical separases have also been found in several species. For example, *Drosophila* separase consists of two proteins, THR and SSE, which correspond to the N- and C-terminal regions of single-chain separase (Jager et al. 2001).

The N-terminal region is mainly composed of helices and was predicted to be ARM or HEAT repeats (Fig. 4.1) (Viadiu et al. 2005). For example, human separase was reported to contain 26 ARM repeats and is responsible for securin binding. This region was also predicted to have a super-helical structure like TPR repeats (Winter et al. 2015). Structural studies of separase and its complex with securin from yeast (*S. cerevisiae*) and *C. elegans*, however, show that it is a mixture of both (Luo and Tong 2017; Boland et al. 2017).

The C-terminal region was originally predicted to have two consecutive caspase domains. The structural studies reveal that this region is composed of two domains, a substrate-binding domain (SD) or pseudo protease domain (PPD), followed by the catalytic domain (CD) or the active protease domain (APD) (Luo and Tong 2017; Lin et al. 2016). Only the CD/APD has activity and structural similarity to caspases.

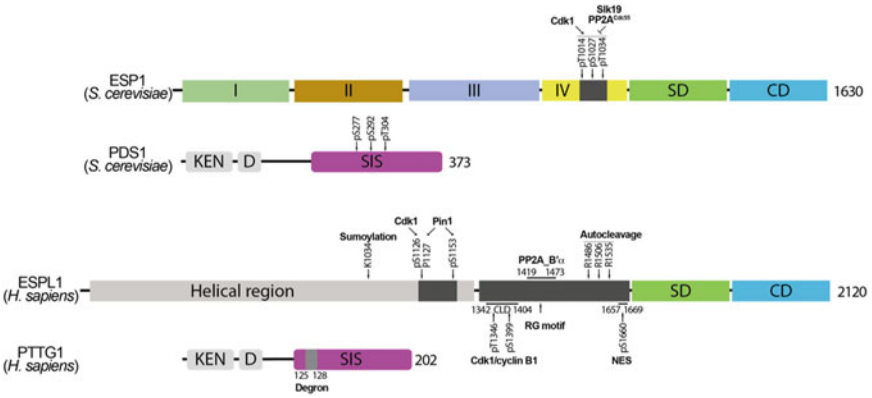


Fig. 4.1 Domain organization of budding yeast and human separase and securin. Domains are depicted as color-coded boxes and labeled. The regulatory segments are shown in dark gray. Modification sites, motifs, and interacting factors are also labeled. The N-terminal helical region of human separase without high-resolution structural information is colored gray

The CD contains a catalytic dyad of a histidine and a cysteine, and the substrate recognition is ensured by a few conserved residues in this domain.

Between the N- and C-terminal regions of separase is a natively unstructured segment with variable length, which contains motifs for binding other proteins and modification (Fig. 4.1). In higher eukaryotes and yeast, this region contains residues that are phosphorylated by Cdk1, the master regulatory kinase of mitosis (Stemmann et al. 2001; Lianga et al. 2018). A Cdc6-like domain (CLD) in this segment is also phosphorylated and crucial for mutual inhibition of the Cdk1-cyclin B1 complex (Boos et al. 2008; Gorr et al. 2005). In addition, several auto-cleavage sites are located in this segment (Fig. 4.1) (Waizenegger et al. 2002), and their cleavage is important for mitosis progression (Papi et al. 2005) and the interaction with PP2A (Holland et al. 2007). Phosphorylation-dependent isomerization mediated by the peptidyl-prolyl isomerase Pin1 on the Ser-Pro motifs of this segment is crucial for separase regulation (Hellmuth et al. 2015b). A nuclear export sequence (NES) resides in this segment (Hellmuth et al. 2018; Sun et al. 2006), phosphorylation of which is crucial for separase's role in DNA damage repair (Hellmuth et al. 2018; Sun et al. 2006). Methylation of a RG-motif within the regulatory segment and sumoylation of the Lys1034 residue is also important for its function in DNA damage response (Hellmuth et al. 2018). A cysteine motif (CxCxxC) was recently identified in this segment of metazoan separases (Melesse et al. 2018).

Diverse Functions of Separase

Although the substrates for separase identified so far are limited to only a few proteins, the functions of separase are very diverse. They can be roughly divided into two categories, depending on whether the function requires its catalytic activity. The cohesin subunits Scc1 and its meiosis-specific counterpart Rec8 are the well-known substrates targeted by separase for spatial and temporal chromosome segregation during mitosis and meiosis, respectively (Uhlmann et al. 1999; Buonomo et al. 2000). Besides chromosome segregation, the catalytic activity is also required for DNA damage repair at interphase in yeast (Nagao et al. 2004; McAleenan et al. 2013) and separase is recruited to double-strand breaks (DSBs) to support homology-directed repair (HDR) in mammals (Hellmuth et al. 2018).

The kinetochore and spindle-associated protein Slk19 is another substrate of separase, discovered soon after Scc1. Unlike the Scc1 cohesin subunit, which is subjected to destruction, the cleavage product of Slk19 stabilizes the anaphase spindle (Sullivan et al. 2001) and regulate pericentric cohesion and anaphase onset together with PP2A by counteracting phosphorylation by Cdk1 (Lianga et al. 2018). Separase is localized to spindle pole bodies and spindle midzone via interactions with its inhibitor securin and the protease activity is required for spindle elongation (Jensen et al. 2001; Baskerville et al. 2008).

Kendrin or pericentric protein, a giant coiled-coil centrosome protein crucial for centriole engagement, is specifically cleaved by separase during mitosis which

ensures a timely centriole disengagement at late mitosis to license centriole duplication during S phase (Matsuo et al. 2012; Lee and Rhee 2012; Schockel et al. 2011; Tsou et al. 2009; Nakamura et al. 2009). Separase is responsible for the cleavage of the N-tail of the CENP-A related protein CPAR-1 at the meiosis I anaphase in *C. elegans* and plays a role in the centromere activity (Monen et al. 2015).

Independent of its protease activity, separase takes part in other important cellular events. Separase contributes to mitotic exit, which requires three pathways including the Cdc14 early anaphase release (FEAR) pathway, Cdk inactivation and mitotic exit network (MEN). Separase together with Slk19 play roles in the FEAR pathway to promote the release of Cdc14, a critical Cdk counteracting phosphatase required for mitotic exit, from the nucleolus to enable mitotic exit (Sullivan and Uhlmann 2003; Stegmeier et al. 2002). Separase also directly binds to the Cdc55 subunit of PP2A to suppress its phosphatase activity and facilitate Net1 phosphorylation in the FEAR pathway to initiate mitotic exit (Queralt et al. 2006). On the other hand, separase is phosphorylated by Cdk1 and its catalytic activity is further controlled by forming a stable complex with Cdk1 via its regulatory subunit cyclin B1. Interestingly, this interaction also enables separase as a Cdk1 inhibitor to block its kinase activity to regulate mitotic exit (Gorr et al. 2005).

Through a genome-wide screen for protein interactions, separase was revealed to be related to genes involved in different functions including transposition, DNA repair, sister chromatid segregation and stress response regulation (Ho et al. 2015). Among them, Ty1 integrase was identified as a separase binding partner in different stages of the cell cycle (Ho et al. 2015). This Ty1 integrase-separase interaction functions to remove cohesin and target Ty1 integrase into the genome for retrotransposition.

The diverse functional roles of separase also include membrane trafficking, distinct from cell division, as separase is located on the vesicle membranes of meiosis-specialized cortical granules (CGs) in *C. elegans* (Bembenek et al. 2007; Bai and Bembenek 2017). In other species including human and *Drosophila*, separase is associated with membranes and plays roles in membrane traffic, protein secretion, and Golgi organization (Bacac et al. 2011; Bard et al. 2006).

Regulation of Separase

Due to the irreversibility of cohesin resolution and other functions, multiple levels of regulation exist to control separase activity tightly. The first and best-known regulator is the inhibitory chaperone securin. To ensure timely activation of separase for chromosome segregation, securin interacts with separase and potently inhibits its proteolytic activity until the onset of anaphase when securin is targeted by the APC/C complex via its N-terminal KEN/D-boxes (Fig. 4.1) and destroyed by the proteasome pathway (Ciosk et al. 1998; Zou et al. 1999; Kumada et al. 1998). Studies also revealed that securin inhibits separase by blocking the access of substrates into the active site of separase (Waizenegger et al. 2002) and this was further supported

by an amino acid substitution on securin that turned it into a substrate (Nagao and Yanagida 2006). This association is conserved in eukaryotes including fungi, plants and animals (Cromer et al. 2019). On the other hand, securin functions as a chaperone in a positive way to help solubilize and stabilize separase (Hornig et al. 2002). Securin binds to nascent separase when it reaches a size of 627 amino acids to assist its proper folding (Hellmuth et al. 2015a). Moreover, in yeast, securin guides the nuclear localization of separase (Agarwal and Cohen-Fix 2002; Jensen et al. 2001) and mediates the proper loading onto the spindle for accelerating spindle elongation (Jensen et al. 2001). This localization process depends on the phosphorylation of three residues of securin (Ser277/Ser292/Thr304) by Cdc28/Cdk1 which also enhances the interaction with separase (Agarwal and Cohen-Fix 2002). Under replication stress conditions, this complex formation and spindle elongation can be repressed through dephosphorylation by PP2A-Cdc55, a key phosphatase of cell cycle progression (Khondker et al. 2019).

Another widely recognized regulator in higher eukaryotes is the Cdk1-cyclin B1 complex, which controls separase activity in both a positive and negative manner like securin. Phosphorylation of separase by Cdk1 was first observed in frog as an additional way to curb its activity independent of securin (Stemmann et al. 2001). Ser1126 of human separase is phosphorylated during metaphase and is dephosphorylated upon anaphase onset for activation. Further investigation showed that this inhibition could sufficiently compensate the inhibition by securin, and it is achieved by cyclin B1 binding to the CLD of separase after the Thr1346/Ser1399 residues in this domain are phosphorylated (Gorr et al. 2005; Boos et al. 2008). This observation is also confirmed by the fact that a normally executed mitosis was seen in securin knock-out human cells (Pfleghaar et al. 2005). Similar to securin, cyclin B1 is also a target of the APC/C machinery (King et al. 1995; Hagting et al. 2002) and its destruction is critical for separase activation during the transition of metaphase to anaphase. Moreover, securin and Cdk1-cyclin B1 interact with separase in a mutually exclusive manner (Gorr et al. 2005). Intriguingly, Ser1126 phosphorylation causes destabilization and aggregation of separase leading to an inactive state. Resembling securin, the Cdk1-cyclin B1 complex associates with separase to stabilize but inhibit the activity until the onset of anaphase (Hellmuth et al. 2015a). Recently, Cdk1-cyclin B2 complex was reported to inhibit separase activity in meiosis I in mouse oocyte (Li et al. 2019).

Another regulatory factor that has direct interactions with separase is the phosphatase PP2A, a central player of cell cycle progression. In mammals, a 55-amino acid segment next to the auto-cleavage sites in the regulatory segment of separase forms a complex with a specific subtype of heterotrimeric PP2A through its regulatory subunit B56 (Holland et al. 2007). Interestingly, distinct from the mutually exclusive manner in the interaction of separase with securin and cyclin B1, PP2A can form a heterotrimeric complex with separase and securin but not Cdk1-cyclin B1. This interaction is abolished by the auto-cleavage in the regulatory segment of human separase (Holland et al. 2007). The bound PP2A can dephosphorylate to stabilize separase-associated securin, thus to delay the APC/C-mediated degradation to regulate separase activity for abrupt chromosome segregation (Hellmuth et al. 2014). In

yeast, separase interacts with and inhibits PP2A containing Cdc55, a different type of regulatory subunit, to allow Scc1 phosphorylation to initiate mitotic exit (Yaakov et al. 2012; Queralt et al. 2006). On the other hand, the dephosphorylation activity of $PP2A^{Cdc55}$ on securin disrupts the separase-securin complex, in which phosphorylation of three sites in securin enhances complex formation (Agarwal and Cohen-Fix 2002) to inhibit spindle elongation during replication stress (Khondker et al. 2019).

Post-translational modifications on separase also control the function of separase at multiple levels. The auto-cleavage of separase is among the first to be elucidated and this modification only exists in vertebrates. Removal of securin allows separase to undergo auto-cleavage at three sites (Arg1486/Arg1506/Arg1535) in the regulatory segment, and the N- and C-terminal cleavage products remain associated with each other (Waizenegger et al. 2002; Zou et al. 2002). Interestingly, the cleavage does not affect its interaction with securin, and the cleavage process is not required for separase activation or the proteolytic activity (Waizenegger et al. 2002; Zou et al. 2002). However, the cleavage was reported to coordinate mitosis entry and progression by affecting the spindle assembly and metaphase chromosome alignment (Papi et al. 2005). Another study showed that the auto-cleavage influences the separase-PP2A complex formation and centromeric cohesion in human cells (Holland et al. 2007).

Phosphorylation and dephosphorylation of separase and its inhibitors and substrates are crucial for the regulation of its other physiological functions. As mentioned above, phosphorylation by Cdk1 on Ser1126 as well as Ser1399 within the CLD in mammals is important for the inhibition (Boos et al. 2008) and stabilization (Hellmuth et al. 2015a) by the Cdk1-cyclin B1 complex. In mouse, this phosphorylation is critical for germ cell development, early embryogenesis (Huang et al. 2009) as well as oogenesis (Xu et al. 2011). Separase in budding yeast is phosphorylated by Cdk1 on the regulatory segment (Thr1014/Ser1027/Thr1034), and this phosphorylation is necessary for separase activation and anaphase onset. PP2A-Cdc55 and Slk19 proteins function along with Cdk1-mediated phosphorylation to inhibit separase activity in addition to securin in yeast (Lianga et al. 2018).

A recent intriguing study revealed another level of regulation of human separase. Phosphorylated human separase is a target of Pin1 which is a phosphorylation-specific peptidyl-prolyl cis/trans isomerase and critical for mitosis regulation (Hellmuth et al. 2015b). Pin1 mediated isomerization on the phosphorylated Ser1126-P1127 motif is required for the binding and inhibition of the Cdk1-cyclin B1 complex. This isomerization relies on Pin1′s docking onto the phosphorylated Ser1153-Pro1154 site via its WW domain (Hellmuth et al. 2015b). This conformational change of separase not only removes the possibility of rebinding by residual securin but also shortens the half-life of separase to ensure a rapid termination of the proteolytic activity for mitosis exit and cohesin reloading (Hellmuth et al. 2015b).

Genome-wide screenings for regulators of separase have also been carried out and reported. In *C. elegans*, separase is crucial for cortical granule exocytosis and cytokinesis. Genetic screening for factors related to this function identified the protein phosphatase 5 (PPH-5) as an extragenic suppressor of separase, which affects the localization of separase on vesicles and suppresses exocytosis and cytokinesis, implying a separase regulatory pathway involving phosphorylation (Richie et al. 2011). A yeast

genetic screen identified new factors that interact with separase and discovered functional relevance of separase in a few cellular processes including retrotransposition and the involved factor Ty1 integrase (Ho et al. 2015).

In chronic myeloid leukemia (CML) patients treated with the TK inhibitor drug Imatinib, the separase protein level is reduced while the proteolytic activity increased in BCR-ABL-positive cells, suggesting a feedback regulation mechanism in tumor cells (Haass et al. 2015b). Further studies showed that separase expression is controlled by the transcription factor c-Myb in CML cell lines (Prinzhorn et al. 2016), indicating separase regulation at the transcriptional level.

Separase-Related Diseases and Drug Development

Given its essential roles in chromosome segregation and DNA damage repair, separase is tightly regulated and its malfunction causes aneuploidy and genome instability leading to severe consequences including cancers in higher vertebrates. Overexpression of separase induces aneuploidy and is often observed in a variety of human cancers including breast, bone, brain, and prostate cancers, thus separase was defined as an oncogene (Pati 2008). For example, separase was reported to be overexpressed in more than 50% of human breast cancers (Mukherjee et al. 2014b; Zhang et al. 2008). The overexpression level of separase highly correlates with tumor progression and reduced survival (Meyer et al. 2009; Pati 2008). Thus, separase is an attractive cancer drug target. Substrate-mimic inhibitors have been reported (Waizenegger et al. 2002). In order to find inhibitors of separase, in vitro (Pfleghaar et al. 2005; Basu et al. 2009) and cell-based (Haass et al. 2015a; Henschke et al. 2019) separase activity assays have been developed. Discovery of small-molecule inhibitors using these assays has been reported (Zhang et al. 2014; Henschke et al. 2019).

Structures of the Separase-Securin Complex

In order to understand how separase functions and how it is regulated at the molecular level, structural information on separase and its complex with securin has been actively pursued since its discovery. Low-resolution negative-stain electron microscopy studies of the human separase-securin complex provided the first glimpse of its overall structure, which adopts an elongated whale-like shape and the blurred tail density reflects its flexible nature (Viadiu et al. 2005). Detailed structural information was not available until the publication of a high-resolution crystal structure of the C-terminal catalytic region of separase in complex with a substrate-mimic peptide inhibitor (Lin et al. 2016), revealing details of the active site and how the substrate is recognized. The structure also showed that, instead of two tandem caspase domains

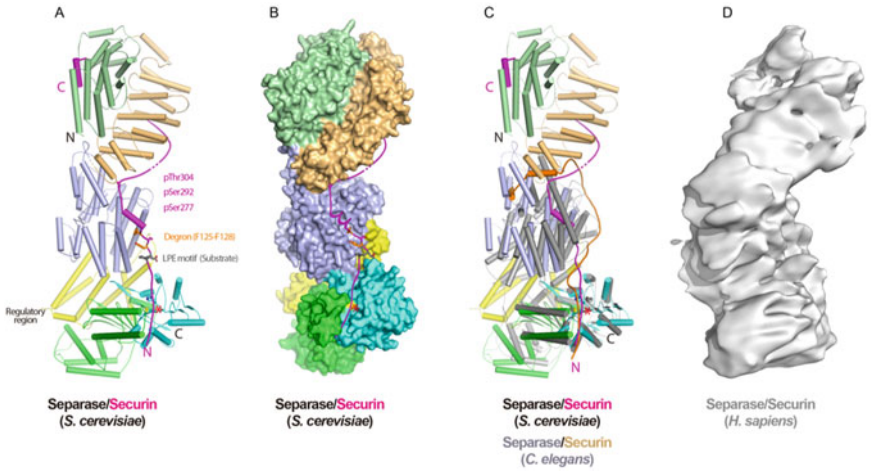


Fig. 4.2 Structures of separase-securin complexes. **a** Budding yeast separase-securin complex structure is represented as a cylindrical cartoon and color-coded as in Fig. 4.1. The active site is labeled with a red star. Modification sites on securin SIS are shown as sticks and are colored and labeled differently. **b** Surface representation in a different view showing the long securin-binding groove in different domains of separase. **c** Superimposition of the structures of yeast and *C. elegans* separase-securin complexes showing that the *C. elegans* separase lacks domains I and II. Colors of different chains are indicated. **d** Low-resolution EM density map of human separase-securin complex (EMDB ID: EMD-3584)

predicted based on sequence analysis, the C-terminal region consists of two subdomains with one resembling the caspase domain and the other sharing no similarity to any known structures (Lin et al. 2016).

Soon after that report, the first high-resolution X-ray crystal structure was published for the near full-length separase-securin complex from budding yeast (Luo and Tong 2018, 2017) (Fig. 4.2a), revealing how the domains of separase are organized and how securin interacts with separase. Yeast separase adopts an elongated conformation with the longest dimension of over 160 Å and it can be divided into six domains. The C-terminal region contains a substrate-binding domain (SD, or PPD) and a catalytic domain (CD, or APD). The structure of this region is similar to the previous structure (Lin et al. 2016) except for the L4 loop in the CD, which forms a two-stranded β-sheet. The N-terminal helical region consists of four domains, designated from I to IV (Figs. 4.1 and 4.2a). Domains I and II have HEAT repeat-like structures and are arranged into a head-to-tail complex. Domain III is a TPR-like super-solenoid structure, quite different from what was predicted as ARM/HEAT repeats. Domain IV is composed of several helices and bridges the helical region and the C-terminal region. The regulatory segment is likely not far from the active site, and could affect the activity upon its modification.

The C-terminal separase-interaction sequence (SIS) of securin (Fig. 4.1) assumes a mostly extended conformation, with only a few ordered secondary structure elements, and interacts with all domains of separase as a chaperone (Fig. 4.2b). The N-terminal

part of SIS acts as a pseudo-substrate and inserts into the substrate-binding site and catalytic site of separase. It is not cleaved by separase as its binding mode in the active site is not compatible with catalysis, with the scissile carbonyl group 6 Å away from the catalytic Cys residue. The C-terminal part binds along separase with the very C-terminus sitting in a cleft formed by domain I, and this observation might explain why securin binds nascent separase only after it reaches a certain length (Hellmuth et al. 2015a). A degron motif in securin (Phe125/Phe128 in humans) or cyclin B1 is only accessible on free securin or cycline B1, to ensure the degradation of excess securin or cyclin B1 proteins through the proteasome pathway in prometaphase to promote timely onset of anaphase (Thomas et al. 2019; Levasseur et al. 2019; Hellmuth et al. 2014). This securin degron motif is buried in the interface with separase in the structure.

The structure of the *C. elegans* separase-securin complex was determined by cryo electron microscopy (Boland et al. 2017). Compared with the complex from other species, this atypical separase is smaller in size owing to a shorter helical region. Superimposed with the yeast separase-securin structure, their C-terminal regions align well and securin is located in the active site in a similar manner (Fig. 4.2c). However, the helical domain lacks the counterparts of domains I and II in the yeast structure, and thus the securin interaction is different in this region. In the same work, a low-resolution cryo-EM map of the human separase-securin complex was also reported. Its shape is reminiscent of the yeast complex, indicating the structural conservation for this important protease through evolution (Fig. 4.2d).

Summary

Separase is a large cysteine protease, with an N-terminal helical region, a C-terminal catalytic region, and an unstructured regulatory segment in between. Separase has diverse functions including chromosome segregation, DNA damage repair, spindle elongation, and centrosome replication. Its activity is tightly controlled with multiple-layer regulation through inhibitors and modifications. Among them, the inhibitory chaperone securin is the most important regulator and the most extensively studied. Securin binds to separase cotranslationally and stabilizes separase until the anaphase onset, when securin is destroyed by the proteasome and separase activity is released. Dysregulation of separase causes genome instability leading to diseases including cancers in humans, making separase a potential target for discovery of inhibitors for cancer therapy. The structural information for the separase-securin complex from several different organisms provides great insights into this enzyme and molecular mechanisms for the inhibition by securin.

However, there are still many unanswered questions and further studies are needed. In human separase, how auto-cleavage, phosphorylation, and isomeric conformational change regulate the activation is still elusive. Phosphorylation of yeast separase is also reported to be critical for activation (Lianga et al. 2018), and how this activation is different from human separase is also an interesting topic for more

research. Protein factors such as PP2A, Cdk1-cyclin B1, Pin1, Slk19 associate with separase mostly through the regulatory segment. Elucidating the molecular details for these interactions and their roles is also critical for us to better understand separase regulation.

The exact recognition mechanism for the C-terminal part of the cleavage site is not known, although a recent study identified an extra binding motif LPE on Scc1 that is similar to a site on securin, which implies the substrate might bind along the securin binding surface (Rosen et al. 2019). The L4 loop in the CD appears to be flexible when alone but ordered in the complex with securin, and the functional role (if any) of this loop needs to be elucidated as well. The additional structural information could also help the discovery of small-molecule inhibitors of separase.

Acknowledgements This research is supported by grant R35GM118093 from the US NIH to LT and a start-up grant of Shanghai Jiao Tong University to SL and a research award from the NSFC (National Natural Science Foundation of China) to SL.

References

Agarwal R, Cohen-Fix O (2002) Phosphorylation of the mitotic regulator Pds1/securin by Cdc28 is required for efficient nuclear localization of Esp1/separase. Genes Dev 16(11):1371–1382. https://doi.org/10.1101/gad.971402

Bacac M, Fusco C, Planche A, Santodomingo J, Demaurex N, Leemann-Zakaryan R, Provero P, Stamenkovic I (2011) Securin and separase modulate membrane traffic by affecting endosomal acidification. Traffic 12(5):615–626. https://doi.org/10.1111/j.1600-0854.2011.01169.x

Bai X, Bembenek JN (2017) Protease dead separase inhibits chromosome segregation and RAB-11 vesicle trafficking. Cell Cycle 16:1902–1917

Bard F, Casano L, Mallabiabarrena A, Wallace E, Saito K, Kitayama H, Guizzunti G, Hu Y, Wendler F, Dasgupta R, Perrimon N, Malhotra V (2006) Functional genomics reveals genes involved in protein secretion and Golgi organization. Nature 439(7076):604–607. https://doi.org/10.1038/nature04377

Baskerville C, Segal M, Reed SI (2008) The protease activity of yeast separase (esp1) is required for anaphase spindle elongation independently of its role in cleavage of cohesin. Genetics 178(4):2361–2372. https://doi.org/10.1534/genetics.107.085308

Basu D, Zhang N, Panigrahi AK, Horton TM, Pati D (2009) Development and validation of a fluorogenic assay to measure separase enzyme activity. Anal Biochem 392(2):133–138. https://doi.org/10.1016/j.ab.2009.05.046

Baum P, Yip C, Goetsch L, Byers B (1988) A yeast gene essential for regulation of spindle pole duplication. Mol Cell Biol 8(12):5386–5397. https://doi.org/10.1128/mcb.8.12.5386

Bembenek JN, Richie CT, Squirrell JM, Campbell JM, Eliceiri KW, Poteryaev D, Spang A, Golden A, White JG (2007) Cortical granule exocytosis in C. elegans is regulated by cell cycle components including separase. Development 134 (21):3837–3848. doi:https://doi.org/10.1242/dev.011361

Boland A, Martin TG, Zhang Z, Yang J, Bai XC, Chang L, Scheres SH, Barford D (2017) Cryo-EM structure of a metazoan separase-securin complex at near-atomic resolution. Nat Struct Mol Biol 24(4):414–418. https://doi.org/10.1038/nsmb.3386

Boos D, Kuffer C, Lenobel R, Korner R, Stemmann O (2008) Phosphorylation-dependent binding of cyclin B1 to a Cdc6-like domain of human separase. the Journal of Biological Chemistry 283(2):816–823. https://doi.org/10.1074/jbc.M706748200

Buonomo SBC, Clyne RK, Fuchs J, Loidl J, Uhlmann F, Nasmyth K (2000) Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by separin. Cell 103(3):387–398. https://doi.org/10.1016/S0092-8674(00)00131-8

Ciosk R, Zachariae W, Michaelis C, Shevchenko A, Mann M, Nasmyth K (1998) An ESP1/PDS1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast. Cell 93(6):1067–1076

Cromer L, Jolivet S, Singh DK, Berthier F, De Winne N, De Jaeger G, Komaki S, Prusicki MA, Schnittger A, Guerois R, Mercier R (2019) Patronus is the elusive plant securin, preventing chromosome separation by antagonizing separase. Proc Natl Acad Sci U S a 116(32):16018–16027. https://doi.org/10.1073/pnas.1906237116

Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters J-M (2019) DNA loop extrusion by human cohesin. Science 366:1338–1345

Gandhi R, Gillespie PJ, Hirano T (2006) Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. Current Biology : CB 16(24):2406–2417. https://doi.org/10.1016/j.cub.2006.10.061

Gimenez-Abian JF, Sumara I, Hirota T, Hauf S, Gerlich D, de la Torre C, Ellenberg J, Peters JM (2004) Regulation of sister chromatid cohesion between chromosome arms. Current Biology : CB 14(13):1187–1193. https://doi.org/10.1016/j.cub.2004.06.052

Gorr IH, Boos D, Stemmann O (2005) Mutual inhibition of separase and Cdk1 by two-step complex formation. Mol Cell 19(1):135–141. https://doi.org/10.1016/j.molcel.2005.05.022

Haass W, Kleiner H, Muller MC, Hofmann WK, Fabarius A, Seifarth W (2015a) Measurement of Separase Proteolytic Activity in Single Living Cells by a Fluorogenic Flow Cytometry Assay. PLoS ONE 10(8):e0133769. https://doi.org/10.1371/journal.pone.0133769

Haass W, Kleiner H, Weiss C, Haferlach C, Schlegelberger B, Muller MC, Hehlmann R, Hofmann WK, Fabarius A, Seifarth W, Schweizerische Arbeitsgemeinschaft fur Klinische K, German CMLSG (2015b) Clonal Evolution and Blast Crisis Correlate with Enhanced Proteolytic Activity of Separase in BCR-ABL b3a2 Fusion Type CML under Imatinib Therapy. PLoS One 10 (6):e0129648. doi:10.1371/journal.pone.0129648

Hagting A, Den Elzen N, Vodermaier HC, Waizenegger IC, Peters JM, Pines J (2002) Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. J Cell Biol 157(7):1125–1137. https://doi.org/10.1083/jcb.200111001

Hauf S, Roitinger E, Koch B, Dittrich CM, Mechtler K, Peters JM (2005) Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. PLoS Biol 3(3):419–432. https://doi.org/ARTNe69

Hellmuth S, Stemmann O (2020) Separase-triggered apoptosis enforces minimal length of mitosis. Nature 580:542–547. https://doi.org/10.1038/s41586-020-2187-y

Hellmuth S, Bottger F, Pan C, Mann M, Stemmann O (2014) PP2A delays APC/C-dependent degradation of separase-associated but not free securin. the EMBO Journal 33(10):1134–1147. https://doi.org/10.1002/embj.201488098

Hellmuth S, Pohlmann C, Brown A, Bottger F, Sprinzl M, Stemmann O (2015a) Positive and Negative Regulation of Vertebrate Separase by Cdk1-Cyclin B1 Might Explain why Securin Is Dispensable. the Journal of Biological Chemistry. https://doi.org/10.1074/jbc.M114.615310

Hellmuth S, Rata S, Brown A, Heidmann S, Novak B, Stemmann O (2015b) Human chromosome segregation involves multi-layered regulation of separase by the peptidyl-prolyl-isomerase pin1. Mol Cell 58(3):495–506. https://doi.org/10.1016/j.molcel.2015.03.025

Hellmuth S, Gutierrez-Caballero C, Llano E, Pendas AM, Stemmann O (2018) Local activation of mammalian separase in interphase promotes double-strand break repair and prevents oncogenic transformation. the EMBO Journal. https://doi.org/10.15252/embj.201899184

Hellmuth S, Gomez-H L, Pendas AM, Stemmann O (2020) Securin-independent regulation of separase by checkpoint-induced shugoshin–MAD2. Nature 580:536–541. https://doi.org/10.1038/s41586-020-2182-3

Henschke L, Frese M, Hellmuth S, Marx A, Stemmann O, Mayer TU (2019) Identification of Bioactive Small Molecule Inhibitors of Separase. ACS Chem Biol 14(10):2155–2159. https://doi.org/10.1021/acschembio.9b00661

Holland AJ, Bottger F, Stemmann O, Taylor SS (2007) Protein phosphatase 2A and separase form a complex regulated by separase autocleavage. J Biol Chem 282(34):24623–24632. https://doi.org/10.1074/jbc.M702545200

Hornig NC, Knowles PP, McDonald NQ, Uhlmann F (2002) The dual mechanism of separase regulation by securin. Current Biology : CB 12(12):973–982

Ho KL, Ma L, Cheung S, Manhas S, Fang N, Wang K, Young B, Loewen C, Mayor T, Measday V (2015) A role for the budding yeast separase, Esp1, in Ty1 element retrotransposition. PLoS Genet 11(3):e1005109. https://doi.org/10.1371/journal.pgen.1005109

Huang X, Andreu-Vieyra CV, Wang M, Cooney AJ, Matzuk MM, Zhang P (2009) Preimplantation mouse embryos depend on inhibitory phosphorylation of separase to prevent chromosome missegregation. Mol Cell Biol 29(6):1498–1505. https://doi.org/10.1128/MCB.01778-08

Jager H, Herzig A, Lehner CF, Heidmann S (2001) Drosophila separase is required for sister chromatid separation and binds to PIM and THR. Genes Dev 15(19):2572–2584. https://doi.org/10.1101/gad.207301

Jensen S, Segal M, Clarke DJ, Reed SI (2001) A novel role of the budding yeast separin Esp1 in anaphase spindle elongation: evidence that proper spindle association of Esp1 is regulated by Pds1. J Cell Biol 152(1):27–40. https://doi.org/10.1083/jcb.152.1.27

Karasu ME, Bouftas N, Keeney S, Wassmann K (2019) Cyclin B3 promotes anaphase I onset in oocyte meiosis. J Cell Biol 218(4):1265–1281. https://doi.org/10.1083/jcb.201808091

Khondker S, Kajjo S, Chandler-Brown D, Skotheim J, Rudner A, Ikui A (2019) PP2A^{Cdc55} dephosphorylates Pds1 to inhibit spindle elongation. bioRxiv

Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H (2019) Human cohesin compacts DNA by loop extrusion. Science 366:1345–1349

King RW, Peters JM, Tugendreich S, Rolfe M, Hieter P, Kirschner MW (1995) A 20s Complex Containing Cdc27 and Cdc16 Catalyzes the Mitosis-Specific Conjugation of Ubiquitin to Cyclin-B. Cell 81(2):279–288. https://doi.org/10.1016/0092-8674(95)90338-0

Kueng S, Hegemann B, Peters BH, Lipp JJ, Schleiffer A, Mechtler K, Peters JM (2006) Wapl controls the dynamic association of cohesin with chromatin. Cell 127(5):955–967. https://doi.org/10.1016/j.cell.2006.09.040

Kumada K, Nakamura T, Nagao K, Funabiki H, Nakagawa T, Yanagida M (1998) Cut1 is loaded onto the spindle by binding to Cut2 and promotes anaphase spindle movement upon Cut2 proteolysis. Current Biology : CB 8(11):633–641. https://doi.org/10.1016/s0960-9822(98)70250-7

Lee K, Rhee K (2012) Separase-dependent cleavage of pericentrin B is necessary and sufficient for centriole disengagement during mitosis. Cell Cycle 11(13):2476–2485. https://doi.org/10.4161/cc.20878

Levasseur MD, Thomas C, Davies OR, Higgins JMG, Madgwick S (2019) Aneuploidy in Oocytes Is Prevented by Sustained CDK1 Activity through Degron Masking in Cyclin B1. Dev Cell 48 (5):672–684 e675. doi:https://doi.org/10.1016/j.devcel.2019.01.008

Lianga N, Dore C, Kennedy EK, Yeh E, Williams EC, Fortinez CM, Wang A, Bloom KS, Rudner AD (2018) Cdk1 phosphorylation of Esp1/Separase functions with PP2A and Slk19 to regulate pericentric Cohesin and anaphase onset. PLoS Genet 14(3):e1007029. https://doi.org/10.1371/journal.pgen.1007029

Lin Z, Luo X, Yu H (2016) Structural basis of cohesin cleavage by separase. Nature 532(7597):131–134. https://doi.org/10.1038/nature17402

Li J, Ouyang YC, Zhang CH, Qian WP, Sun QY (2019) The cyclin B2/CDK1 complex inhibits separase activity in mouse oocyte meiosis I. Development 146 (23). doi:https://doi.org/10.1242/dev.182519

Luo S, Tong L (2018) Structural biology of the separase-securin complex with crucial roles in chromosome segregation. Curr Opin Struct Biol 49:114–122. https://doi.org/10.1016/j.sbi.2018.01.012

Luo S, Tong L (2017) Molecular mechanism for the regulation of yeast separase by securin. Nature 542(7640):255–259. https://doi.org/10.1038/nature21061

Matsuo K, Ohsumi K, Iwabuchi M, Kawamata T, Ono Y, Takahashi M (2012) Kendrin is a novel substrate for separase involved in the licensing of centriole duplication. Current Biology : CB 22(10):915–921. https://doi.org/10.1016/j.cub.2012.03.048

May GS, McGoldrick CA, Holt CL, Denison SH (1992) The bimB3 mutation of Aspergillus nidulans uncouples DNA replication from the completion of mitosis. the Journal of Biological Chemistry 267(22):15737–15743

McAleenan A, Clemente-Blanco A, Cordon-Preciado V, Sen N, Esteras M, Jarmuz A, Aragon L (2013) Post-replicative repair involves separase-dependent removal of the kleisin subunit of cohesin. Nature 493 (7431):250-254. https://doi.org/10.1038/nature11630

Melesse M, Bembenek JN, Zhulin IB (2018) Conservation of the separase regulatory domain. Biol Direct 13:7

Meyer R, Fofanov V, Panigrahi A, Merchant F, Zhang N, Pati D (2009) Overexpression and mislocalization of the chromosomal segregation protein separase in multiple human cancers. Clin Cancer Res 15(8):2703–2710. https://doi.org/10.1158/1078-0432.CCR-08-2454

Monen J, Hattersley N, Muroyama A, Stevens D, Oegema K, Desai A (2015) Separase Cleaves the N-Tail of the CENP-A Related Protein CPAR-1 at the Meiosis I Metaphase-Anaphase Transition in C. elegans. PLoS One 10 (4):e0125382. doi:https://doi.org/10.1371/journal.pone.0125382

Mukherjee M, Byrd T, Brawley VS, Bielamowicz K, Li XN, Merchant F, Maitra S, Sumazin P, Fuller G, Kew Y, Sun D, Powell SZ, Ahmed N, Zhang N, Pati D (2014a) Overexpression and constitutive nuclear localization of cohesin protease Separase protein correlates with high incidence of relapse and reduced overall survival in glioblastoma multiforme. J Neurooncol 119(1):27–35. https://doi.org/10.1007/s11060-014-1458-6

Mukherjee M, Ge G, Zhang N, Edwards DG, Sumazin P, Sharan SK, Rao PH, Medina D, Pati D (2014b) MMTV-Espl1 transgenic mice develop aneuploid, estrogen receptor alpha (ERalpha)-positive mammary adenocarcinomas. Oncogene 33(48):5511–5522. https://doi.org/10.1038/onc.2013.493

Nagao K, Yanagida M (2006) Securin can have a separase cleavage site by substitution mutations in the domain required for stabilization and inhibition of separase. Genes Cells 11(3):247–260. https://doi.org/10.1111/j.1365-2443.2006.00941.x

Nagao K, Adachi Y, Yanagida M (2004) Separase-mediated cleavage of cohesin at interphase is required for DNA repair. Nature 430(7003):1044–1048. https://doi.org/10.1038/nature02803

Nakamura A, Arai H, Fujita N (2009) Centrosomal Aki1 and cohesin function in separase-regulated centriole disengagement. J Cell Biol 187(5):607–614. https://doi.org/10.1083/jcb.200906019

Papi M, Berdougo E, Randall CL, Ganguly S, Jallepalli PV (2005) Multiple roles for separase auto-cleavage during the G2/M transition. Nat Cell Biol 7(10):1029–1035. https://doi.org/10.1038/ncb1303

Pati D (2008) Oncogenic activity of separase. Cell Cycle 7(22):3481–3482. https://doi.org/10.4161/cc.7.22.7048

Pfleghaar K, Heubes S, Cox J, Stemmann O, Speicher MR (2005) Securin is not required for chromosomal stability in human cells. PLoS Biol 3(12):e416. https://doi.org/10.1371/journal.pbio.0030416

Prinzhorn W, Stehle M, Kleiner H, Ruppenthal S, Muller MC, Hofmann WK, Fabarius A, Seifarth W (2016) c-MYB is a transcriptional regulator of ESPL1/Separase in BCR-ABL-positive chronic myeloid leukemia. Biomark Res 4:5. https://doi.org/10.1186/s40364-016-0059-2

Queralt E, Lehane C, Novak B, Uhlmann F (2006) Downregulation of PP2A(Cdc55) phosphatase by separase initiates mitotic exit in budding yeast. Cell 125(4):719–732. https://doi.org/10.1016/j.cell.2006.03.038

Richie CT, Bembenek JN, Chestnut B, Furuta T, Schumacher JM, Wallenfang M, Golden A (2011) Protein phosphatase 5 is a negative regulator of separase function during cortical granule exocytosis in C. elegans. J Cell Sci 124 (Pt 17):2903–2913. doi:https://doi.org/10.1242/jcs.073379

Rosen LE, Klebba JE, Asfaha JB, Ghent CM, Campbell MG, Cheng Y, Morgan DO (2019) Cohesin cleavage by separase is enhanced by a substrate motif distinct from the cleavage site. Nat Commun 10(1):5189. https://doi.org/10.1038/s41467-019-13209-y

Schockel L, Mockel M, Mayer B, Boos D, Stemmann O (2011) Cleavage of cohesin rings coordinates the separation of centrioles and chromatids. Nat Cell Biol 13(8):966–972. https://doi.org/10.1038/ncb2280

Stegmeier F, Visintin R, Amon A (2002) Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. Cell 108(2):207–220. https://doi.org/10.1016/s0092-8674(02)00618-9

Stemmann O, Zou H, Gerber SA, Gygi SP, Kirschner MW (2001) Dual inhibition of sister chromatid separation at metaphase. Cell 107(6):715–726

Sullivan M, Uhlmann F (2003) A non-proteolytic function of separase links the onset of anaphase to mitotic exit. Nat Cell Biol 5(3):249–254. https://doi.org/10.1038/ncb940

Sullivan M, Lehane C, Uhlmann F (2001) Orchestrating anaphase and mitotic exit: separase cleavage and localization of Slk19. Nat Cell Biol 3(9):771–777. https://doi.org/10.1038/ncb0901-771

Sumara I, Vorlaufer E, Stukenberg PT, Kelm O, Redemann N, Nigg EA, Peters JM (2002) The dissociation of cohesin from chromosomes in prophase is regulated by polo-like kinase. Mol Cell 9(3):515–525. https://doi.org/10.1016/S1097-2765(02)00473-2

Sun Y, Yu H, Zou H (2006) Nuclear exclusion of separase prevents cohesin cleavage in interphase cells. Cell Cycle 5(21):2537–2542. https://doi.org/10.4161/cc.5.21.3407

Thomas C, Levasseur MD, Harris RJ, Davies OR, Madgwick S (2019) Synchronous chromosome segregation in mouse oocytes is ensured by biphasic securin destruction and cyclin B1-Cdk1. bioRxiv

Tsou MF, Wang WJ, George KA, Uryu K, Stearns T, Jallepalli PV (2009) Polo kinase and separase regulate the mitotic licensing of centriole duplication in human cells. Dev Cell 17(3):344–354. https://doi.org/10.1016/j.devcel.2009.07.015

Uhlmann F, Lottspeich F, Nasmyth K (1999) Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature 400(6739):37–42. https://doi.org/10.1038/21831

Uhlmann F, Wernic D, Poupart MA, Koonin EV, Nasmyth K (2000) Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell 103(3):375–386

Uzawa S, Samejima I, Hirano T, Tanaka K, Yanagida M (1990) The fission yeast cut1+ gene regulates spindle pole body duplication and has homology to the budding yeast ESP1 gene. Cell 62(5):913–925

Viadiu H, Stemmann O, Kirschner MW, Walz T (2005) Domain structure of separase and its binding to securin as determined by EM. Nat Struct Mol Biol 12(6):552–553. https://doi.org/10.1038/nsmb935

Waizenegger IC, Hauf S, Meinke A, Peters JM (2000) Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell 103(3):399–410

Waizenegger I, Gimenez-Abian JF, Wernic D, Peters JM (2002) Regulation of human separase by securin binding and autocleavage. Current Biology : CB 12(16):1368–1378

Winter A, Schmid R, Bayliss R (2015) Structural Insights into Separase Architecture and Substrate Recognition through Computational Modelling of Caspase-Like and Death Domains. PLoS Comput Biol 11(10):e1004548. https://doi.org/10.1371/journal.pcbi.1004548

Xu J, Wang M, Gao X, Hu B, Du Y, Zhou J, Tian X, Huang X (2011) Separase phosphosite mutation leads to genome instability and primordial germ cell depletion during oogenesis. PLoS ONE 6(4):e18763. https://doi.org/10.1371/journal.pone.0018763

Yaakov G, Thorn K, Morgan DO (2012) Separase Biosensor Reveals that Cohesin Cleavage Timing Depends on Phosphatase PP2A(Cdc55) Regulation. Dev Cell 23(1):124–136. https://doi.org/10.1016/j.devcel.2012.06.007

Zhang N, Ge G, Meyer R, Sethi S, Basu D, Pradhan S, Zhao YJ, Li XN, Cai WW, El-Naggar AK, Baladandayuthapani V, Kittrell FS, Rao PH, Medina D, Pati D (2008) Overexpression of Separase

induces aneuploidy and mammary tumorigenesis. Proc Natl Acad Sci U S A 105 (35):13033-13038. doi:10.1073/pnas.0801610105

Zhang NG, Scorsone K, Ge GQ, Kaffes CC, Dobrolecki LE, Mukherjee M, Lewis MT, Berg S, Stephan CC, Pati D (2014) Identification and Characterization of Separase Inhibitors (Sepins) for Cancer Therapy. J Biomol Screen 19(6):878–889. https://doi.org/10.1177/1087057114520972

Zou H, McGarry TJ, Bernal T, Kirschner MW (1999) Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. Science 285(5426):418–422

Zou H, Stemmann O, Anderson JS, Mann M, Kirschner MW (2002) Anaphase specific auto-cleavage of separase. FEBS Lett 528(1–3):246–250

Chapter 5
The DNA Replication Machine: Structure and Dynamic Function

Nina Y. Yao and Michael E. O'Donnell

Abstract In all cell types, a multi-protein machinery is required to accurately duplicate the large duplex DNA genome. This central life process requires five core replisome factors in all cellular life forms studied thus far. Unexpectedly, three of the five core replisome factors have no common ancestor between bacteria and eukaryotes. Accordingly, the replisome machines of bacteria and eukaryotes have important distinctions in the way that they are organized and function. This chapter outlines the major replication proteins that perform DNA duplication at replication forks, with particular attention to differences and similarities in the strategies used by eukaryotes and bacteria.

Keywords Replisome · Helicase · DNA polymerase · Primase · Sliding clamp · Clamp loader

Introduction

Cellular DNA genomes are duplicated by a multiprotein machinery referred to as a replisome (Kornberg and Baker 1992). Replisomes of bacteria, archaea and eukaryotes utilize a core set of five distinctive activities that we refer to here as the "core" replisome factors (Bell and Labib 2016; Yao and O'Donnell 2019). The core replisome factors of Escherichia coli and Saccharomyces cerevisiae, representative of bacteria and eukaryotes respectively, are listed in Table 5.1. The core replisome factors are: (1) a hexameric helicase, that encircles one strand of DNA and uses ATP

N. Y. Yao · M. E. O'Donnell (✉)
DNA Replication Laboratory, The Rockefeller University, New York, USA 10065
e-mail: odonnel@rockefeller.edu

N. Y. Yao
e-mail: Nina.Yao@rockefeller.edu

M. E. O'Donnell
Howard Hughes Medical Institute, The Rockefeller University, New York, USA 10065

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_5

Table 5.1 The five core replisome factors of bacteria and eukaryotes

Cell type	Helicase	Primase	DNA polymerase	Clamp	Clamp loader
Bacteria (E. coli)	DnaB (Homohexamer)	Primase (Monomer)	Pol III (α Pol; ε exo; θ)	Beta homodimer	Tau complex $\tau_3\delta\delta'\chi\psi$
Eukaryote (S. cerevisiae)	CMG: Cdc45, Mcm 2, 3, 4, 5, 6, 7, and GINS (Psf1, 2, 3, and Sld5)	Pol α-primase (Pol 1, Pol 12, Pri 1, Pri 2)	Pol ε: Pol 2 (Pol 2, Dpb2, Dpb3, Dpb4) Pol δ: Pol 3, Pol 31, Pol 32	PCNA (homotrimer)	RFC (RFC 1,2,3,4,5)

to motor along the ssDNA and unwind the duplex, (2) DNA polymerases, which polymerize DNA and contain a 3'–5' proofreading exonuclease, (3) a primase that synthesizes RNA (or RNA-DNA) primers to initiate elongation by DNA polymerases, (4) a sliding clamp ring that encircles double-strand (ds) DNA (or RNA-DNA) and tethers DNA polymerases to a primed DNA for highly processive synthesis, and (5) a clamp loader complex that uses ATP to open and close the circular sliding clamp around DNA. There are two other replication proteins that are important to genome duplication in all cell types but are not always considered part of the replisome machine. One of these replication proteins is the single-strand (ss) DNA binding protein (e.g. bacterial SSB (Single-Strand Binding protein) and eukaryotic RPA (Replication Protein A)) that coat ssDNA to melt hairpin blocks to polymerization and also to protect ssDNA from nucleases. The other replication protein is a DNA topoisomerase(s) that removes helical turns from dsDNA, needed for progression of the replisome. DNA topoisomerases are thought to act separate (and ahead) of replication forks. The archaeal replisome is quite similar to the eukaryotic replisome as the archaeal replication machinery is evolutionarily related to the eukaryotic replisome, and interested readers are referred to reviews on archaeal replication (Makarova et al. 2012; Makarova et al. 2014).

A comparison of the subunit composition of the core replisome factors of bacteria (E. coli) and eukaryotes (S. cerevisiae) shows that the eukaryotic core factors are multi-subunit complexes giving eukaryotic replisomes many more distinct polypeptides compared to bacterial replisomes (Table 5.1). For example, the eukaryotic replicative helicase is composed of 11 subunits and is referred to as CMG for its 3 subcomponents, the Cdc45 (Cell Division Cycle 45) protein, the Mcm (Mini Chromosome Maintenance) 2–7 heterohexamer motor, and the 4 subunit GINS complex (Psf 1,2,3 and Sld 5) (Ilves et al. 2010; Moyer et al. 2006). In contrast, the DnaB helicase of the E. coli replisome is a single polypeptide that forms a homohexamer (Kornberg and Baker 1992; LeBowitz and McMacken 1986). The eukaryotic primase is the four-subunit DNA polymerase alpha-primase that contains a DNA polymerase (Pol 1) and RNA primase (Pri 1/2) in addition to the Pol 12 subunit and generates a RNA-DNA primer of about 25 nucleotides (Conaway and Lehman 1982). In contrast, the E. coli primase is only a single subunit that makes RNA primers of 10-12

nucleotides in length (Kornberg and Baker 1992). The eukaryotic replisome utilizes two distinct DNA polymerases, epsilon and delta, for replication of the leading and lagging strands, respectively; each are multisubunit complexes (Burgers and Kunkel 2017). The E. coli replisome utilizes identical DNA polymerase III enzymes for both strands, and while the E. coli polymerase III is composed of separate DNA polymerase and 3'–5' exonuclease subunits, the corresponding replicative DNA polymerase of most bacteria is a single subunit that contains both polymerase and exonuclease activities in the same polypeptide chain (Kornberg and Baker 1992; McHenry 2011). The sliding clamp of eukaryotes is the homotrimeric PCNA (Proliferating Cell Nuclear Antigen), and in bacteria it is the dimeric beta, but each form rings with similar dimensions and molecular weight (Itsathitphaisarn et al. 2012; Kong et al. 1992). The eukaryotic clamp loader, RFC (Replication Factor C), is a heteropentamer of proteins that each share homology to AAA + proteins (ATPases Associated with diverse Activities), and the E. coli clamp loader also contains five AAA + subunits (Jeruzalmi et al. 2001; Kelch et al. 2012). The E. coli SSB is a homotetramer and each subunit contains an OB (Oligosaccharide Binding) fold, while eukaryotic RPA is a heterotrimer that also uses four OB folds to bind ssDNA (Antony and Lohman 2019; Chen and Wold 2014). The eukaryotic replisome contains several "accessory" factors that travel with the replisome and have no functional counterpart in bacteria (Gambus et al. 2006; Gambus et al. 2009). Although we focus on the core replisome factors is this review, we discuss a few of the eukaryotic replisome accessory factors that have significant impact on undisturbed replisome action. The known eukaryotic replisome accessory factors include: the Ctf4 (Chromosome Transmission Fidelity) homotrimer (metazoan AND1), the Mcm10 (Mini Chromosome Maintenance) initiation factor, the Mrc1 (Mediator of Replication Checkpoint) mediator of the S-phase DNA checkpoint response (metazoan CLASPIN), the Tof1-Csm3 (Topoisomerase 1 interacting Factor and Chromosome Segregation in Meiosis) paused fork stabilizing subunits (metazoan TIM-TIPPIN), and the FACT (FAcilitates Chromatin Transcription) nucleosome mobility factor (Gambus et al. 2006; Gambus et al. 2009). Of these replisome accessory factors, only Mcm10 is essential to S. cerevisiae cell viability.

Evolution and Structure of the Core Replisome Components

The advent of genomic sequencing revealed the surprising finding that major components of bacterial and eukaryotic replisomes have non-homologous sequences and thus are presumed to have evolved independently (Table 5.2) (Forterre et al. 2004;

Table 5.2 Structural folds in the non-homologous core replisome factors of bacteria and eukaryotes

Component	Bacteria	Eukaryote
Helicase	RecA fold	AAA + fold
Primase	Toprim fold	Pol X fold
DNA Polymerase	Polymerase III	Polymerases ε and δ

Koonin et al. 2015; Yao and O'Donnell 2016). This stands in contrast to other major information pathways in which there exists homology between bacterial and eukaryotic ribosomes and RNA polymerase subunits. The three core replisome factors that lack homology are as follows. Bacterial helicase (e.g. E. coli DnaB) uses a RecA fold for ATPase driven function, while the MCM motors of eukaryotic CMG utilize a AAA + fold for ATP binding and hydrolysis. DNA polymerases assort into distinct non-homologous sequence families (Steitz 1999; Yang and Woodgate 2007), and the eukaryotic leading and lagging strand DNA polymerases epsilon and delta are in the B-family of polymerases, while the bacterial replicative polymerase (Pol III/PolC) is in the C-family of polymerases. The bacterial primase is derived from a topoisomerase fold (Aravind et al. 1998; Podobnik et al. 2000), while the RNA primase in the eukaryotic DNA polymerase alpha-primase has the folding pattern of the DNA polymerase X family (Kilkenny et al. 2013; Kirk and Kuchta 1999).

Helicase

Replicative hexameric ring shaped helicases in both bacteria and eukaryotes are composed of 6 motor subunits that each have a N-terminal and a C-terminal domain (Fig. 5.1a), giving the hexameric ring an appearance of two stacked rings on side view (Fig. 5.1a) (O'Donnell and Li 2018). Thus the hexamer consists of an N-tier ring and a C-tier ring. The ATP binding sites are located in the C-terminal domain in

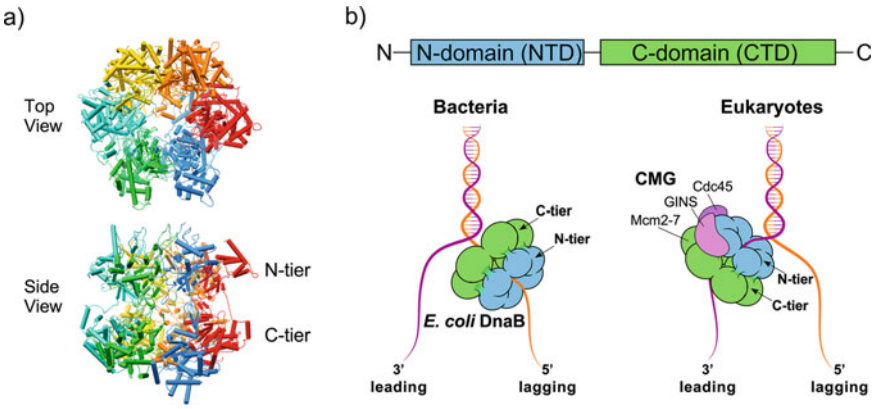


Fig. 5.1 Replicative helicases are hexameric rings with two tiers. **a** Top and side views of the eukaryotic Mcm2-7 heterohexamer of the CMG helicase; the GINS and Cdc45 subunits are omitted for clarity (PDB 3JC7). The side view shows that the two domains of each subunit give the helicase the appearance of two stacked rings, forming a N-tier and a C-tier. **b** Top: Each subunit of replicative hexameric helicases have an N-terminal and C-terminal domain. The ATP sites are located in the C-terminal domain for both bacteria and eukaryotes. Bottom: Hexameric helicases of eukaryotes and bacteria encircle opposite strands because they track in opposite directions on ssDNA. Note also that the eukaryotic helicase tracks N-tier first while the bacterial helicase tracks C-tier first

both E. coli DnaB and the eukaryotic Mcm2–7 motors of CMG, and the ATP sites are positioned at subunit interfaces of the C-tier ring in which residues from both neighboring subunits are required for ATP hydrolytic activity (Enemark and Joshua-Tor 2008; Lyubimov et al. 2011; O'Donnell and Li 2018). The interfacial ATP sites may enable intramolecular communication within the ring during unwinding action. Both the bacterial DnaB and eukaryotic CMG function by steric exclusion, in which they translocate along a ssDNA and act as a wedge at a forked junction to separate the strands; one strand is excluded to the outside of the helicase ring and the other strand is pulled through the central channel of the helicase ring by loops within the motors in the C-terminal domains (Ahnert and Patel 1997; Fu et al. 2011; Hacker and Johnson 1997; Lee et al. 2014). Interestingly, the bacterial and eukaryotic replicative helicases translocate on ssDNA in opposite directions. The RecA based bacterial helicase travels 5'–3', placing it on the lagging strand while AAA + based eukaryotic CMG travels 3'–5', placing it on the leading strand (Fig. 5.1b). Furthermore, the bacterial helicase travels C-tier first (Itsathitphaisarn et al. 2012), while the eukaryotic CMG travels N-tier first (Georgescu et al. 2017). For further information about the detailed structure and mechanism of hexameric helicases the reader is referred to reviews that focus on this subject (Enemark and Joshua-Tor 2008; Lyubimov et al. 2011; O'Donnell and Li 2018).

DNA Polymerase

All known DNA polymerases are shaped as a right hand, having thumb, fingers and palm sub-domains, even though DNA polymerases can be assorted into at least six different families that do not share homology to one another (Steitz 1999; Yang and Woodgate 2007). Many (but not all) DNA polymerases, such as the replicative DNA polymerases, also contain a proofreading 3'–5' exonuclease that removes mismatched nucleotides inserted by mistake. The bacterial replicative polymerase is the sole member of the C-family of DNA polymerases. The eukaryotic DNA polymerases delta and epsilon are in the B-family and thus share homology to one another. The leading strand DNA polymerase epsilon is unique in that it contains two DNA polymerases in its Pol 2 subunit (Tahirov et al. 2009). The N-terminal polymerase domain in Pol 2 of DNA polymerase epsilon is the active DNA polymerase, while the C-terminal polymerase is inactive (Tahirov et al. 2009). The active N-terminal domain of Pol 2 in DNA polymerase epsilon is unique in having an extra sub-domain (P-domain) that joins with the thumb/finger/palm sub-domains to completely encircle the primed template (Hogg et al. 2014). Interestingly, genetic studies reveal that the inactive C-terminal polymerase of Pol 2 is essential, while the N-terminal active polymerase of Pol 2 is not required for cell viability, although the mutant cells grow very slow (Dua et al. 1999; Kesti et al. 1999). It is thought that the catalytic N-terminal half of Pol 2, if deleted, can be substituted by DNA polymerase delta (Dua et al. 1999; Kesti et al. 1999). The essential, but inactive C-terminal polymerase in

Pol 2 is thought to serve an essential structural role, and recent studies indicate that it is important for the assembly of CMG at an origin (Goswami et al. 2018).

Primase

Primases of bacteria and eukaryotes are highly diverged and bear little similarity in sequence or structure. It is still unresolved why the eukaryotic DNA polymerase alpha-primase contains a DNA polymerase and functions with its priming subunits to make a hybrid RNA-DNA primer. Interestingly, a recent study showed that mutating the polymerase active site in Pol 1, the DNA polymerase of polymerase alpha-primase, still gave full replication activity in a single-molecule study, implying RNA primers can be used instead of RNA-DNA hybrid primers (Lewis et al. 2020). However, whether cells require DNA polymerase activity within polymerase alpha-primase for cell viability is not yet known.

Clamps and Clamp Loaders

Unlike the other core factors, the clamp and clamp loaders of bacteria and eukaryotes share the same common ancestor (Kelch et al. 2012; Yao and O'Donnell 2016). The beta and PCNA clamps are essentially superimposable, having the same secondary and tertiary 6 domain structure, with the exception that each monomer of dimeric bacterial beta has three domains (Kong et al. 1992) and each monomer of trimeric PCNA has two domains (Fig. 5.2a) (Gulbis et al. 1996). The clamp loader subunits of bacteria and eukaryotes are heteropentamers that share sequence homology to one another, and display homology from bacteria to eukaryotes (Cullmann et al. 1995; O'Donnell et al. 1993). Clamp loader pentamers form a gapped ring that bind dsDNA in the central chamber (Fig. 5.2b) (Kelch et al. 2011, 2012). Clamp loaders operate by binding/opening their respective clamp and then bind primed DNA in the center of the pentamer ring, thereby positioning the DNA through the clamp (Fig. 5.2b). The DNA induces a conformation change in the clamp loader that activates the ATP sites for hydrolysis, which ejects the clamp loader and enables the clamp to snap shut around DNA. For a more in-depth discussion of the mechanism of clamp loaders, the reader is referred to reviews on the subject (Kelch 2016; Kelch et al. 2012; Yao and O'Donnell 2012).

One may ask why the clamp and clamp loader are conserved while the other core replisome factors have distinct structures and evolutionary heritage? Clamps and clamp loaders are well known to be involved in numerous processes, not just replication (De Biasio and Blanco 2013; Georgescu et al. 2015a; Yao and O'Donnell 2016). Hence, while they were discovered in DNA replication systems, they may have originally evolved for function in other DNA metabolic reactions, and then were recruited later in evolution for DNA replication. For example, use of a clamp

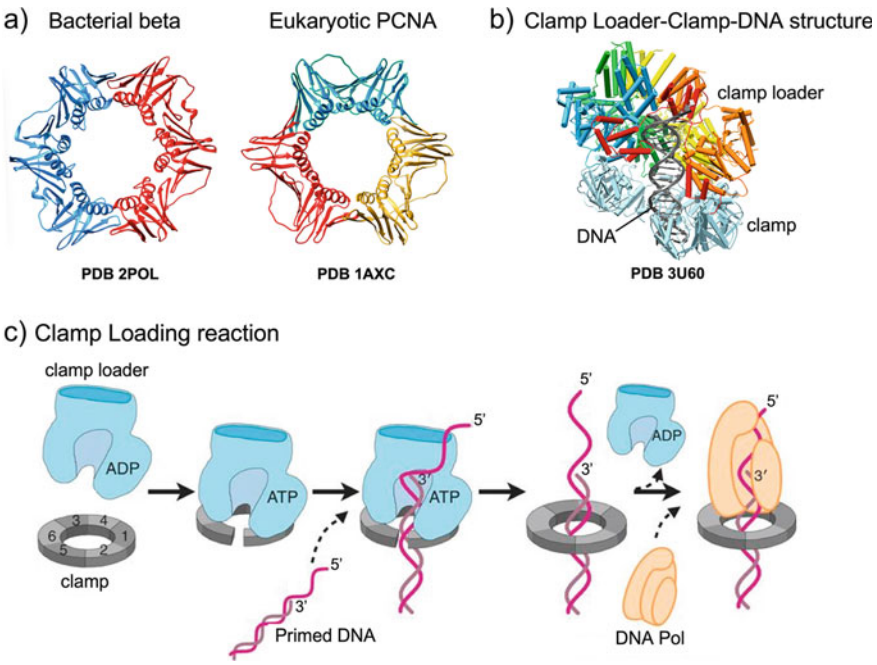


Fig. 5.2 Sliding clamps and clamp loaders. **a** Representative sliding clamps of bacteria (E. coli beta, PDB 2POL) and eukaryotes (human PCNA, PDB 1AXE) are shown in ribbon representation. **b** The clamp loader of T4 bacteriophage bound to its clamp and primed DNA. **c** Steps in the clamp loading reaction. ATP binding to the clamp loader enables it to bind/open the clamp, which then binds primed DNA in the central chamber of the clamp loader, positioning dsDNA through the clamp. ATP hydrolysis ejects the clamp loader, enabling the clamp to close and bind to the DNA polymerase. Panel c is reproduced with permission from Fig. 5.2 in (Jeruzalmi et al. 2001; Kelch et al. 2012)

that binds a variety of different DNA metabolic proteins will enhance their effective concentration on DNA, and thus the cell would not need to produce as much of the protein for its activity on DNA. Specifically, the effective local concentration of a clamp-interacting protein relative to DNA will be greatly enhanced by proximity to DNA via interaction with a mobile sliding clamp that tethers it to DNA.

Considering that bacteria and eukaryotic replisomes evolved separately, why do they have similar core replisome factors? One possible explanation is that genome replication of the Last Universal Common Ancestor (LUCA) cell may have been very different from modern day cells. The LUCA genome could possibly have even been RNA. Viruses and bacteriophage have a plethora of different replication strategies, providing precedent for unique DNA replication processes (Kornberg and Baker 1992). For example, LUCA may have replicated a duplex genome one strand at a time (e.g. first duplicating one strand and then doubling back to replicate the other). This "sequential strand replication" strategy is observed in some dsDNA and RNA viruses today, and can circumvent the need for both primase and helicase.

Primase is circumvented by sequential strand replication by protein priming by phi29 bacteriophage, priming by cleavage of a terminal hairpin by adeno-associated virus, and tRNA priming in retroviruses (Kornberg and Baker 1992). The need for a helicase can be circumvented using a strand displacing DNA polymerase. For example, phi29 bacteriophage replicates its dsDNA genome using a strand displacing polymerase instead of a helicase.

Taking these speculations a step further, it is possible that cells eventually evolved a large enough genome to require a mechanism to rapidly replicate both strands at the same time to be more fit, or competitive with other cells. Rapid and simultaneous replication of both strands of the duplex requires an RNA primase for semi-discontinuous replication, and an ATP driven DNA helicase dedicated to rapid unwinding of the duplex genome. If these requirements for cellular fitness occurred after the split of bacteria and eukaryotes from LUCA, it could possibly explain why bacteria utilize a helicase and primase that share no homology to their eukaryotic counterparts.

Organization of Replisomes

The apparent distinct evolutionary heritage of helicase, primase and DNA polymerases of eukaryotic and bacterial replisomes presents the possibility that these components are organized differently in their respective replisome machines. Thus, we summarize here the current state of knowledge about the organization of the core replisome factors in bacterial and eukaryotic cell types, using E. coli to represent bacteria and S. cerevisiae to represent eukaryotes.

Bacterial Replisome

The bacterial homohexameric replicative helicase (e.g. E. coli DnaB) unwinds DNA 5'–3' and is thus located on the lagging strand (LeBowitz and McMacken 1986). A direct connection of E. coli polymerase III to the replicative DnaB helicase has yet to be observed, and if it exists it is sufficiently weak to avoid detection. However, three of the clamp loading subunits (tau) contain C-terminal appendages that bind both the DnaB helicase and DNA polymerase III (Dallmann et al. 2000; Gao and McHenry 2001; O'Donnell et al. 2001). Thus, the clamp loader is the organizing center of the E. coli replisome, connecting the helicase to the DNA polymerases (Fig. 5.3) (O'Donnell et al. 2001). The primase transiently binds the DnaB helicase for function, and extends primers in the opposite direction of helicase unwinding (Kornberg and Baker 1992). The E. coli primase acts distributively in the replisome, coming on and off the replisome to prime the lagging strand at intervals of 1-2 kb (Tougu and Marians 1996). The lagging strand DNA polymerase III, held to DNA by its beta clamp, is capable of processive synthesis of an Okazaki fragment but

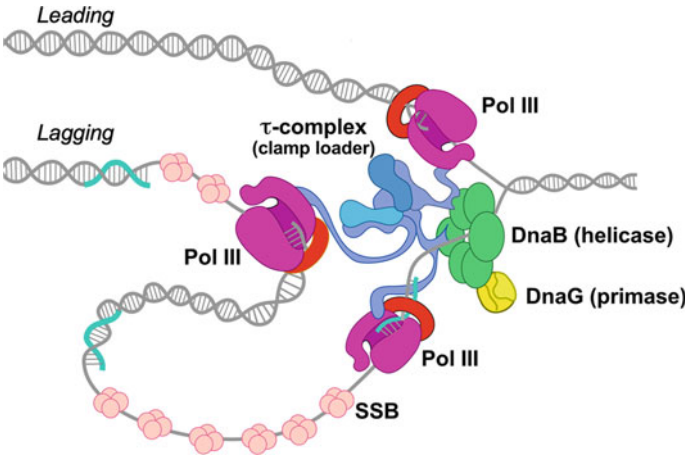


Fig. 5.3 Organization of the core replisome factors of the E. coli replisome. The leading strand DNA polymerase III is held to DNA by the beta clamp. The τ-complex clamp loader contains 3 t subunits, each of which contain extensions that bind a DNA polymerase III molecule and connect to the DnaB helicase hexamer that encircles the lagging strand. Two molecules of DNA polymerase III-beta on the lagging strand. Primase transiently binds the helicase to initiate synthesis of the lagging strand RNA primers onto which the central clamp loader repeatedly loads beta clamps as RNA primers are formed. Extension of a lagging strand primer results in formation of a 1-2 kb Okazaki fragment DNA loop. The SSB tetramer binds the lagging strand ssDNA to protect it from nucleases and to remove secondary structures that would normally impede DNA polymerase movement

has an intrinsic mechanism that ejects DNA polymerase III from the beta clamp and DNA upon completing a segment of DNA, enabling it to recycle to a new RNA primed site while remaining bound to the clamp loader/replisome (O'Donnell 1987; Stukenberg et al. 1994). This polymerase recycling mechanism will be discussed further in the next section. Structural studies of the bacterial cellular replisome have yet to be reported. Hence, specific details about the organization of the replisome factors relative to one another in 3D space is not yet known. However, the structure of the T7 bacteriophage replisome has recently been determined (Gao et al. 2019), and has similarity to the eukaryotic replisome structure, as will be described below.

Considering that leading and lagging strand DNA polymerases are structurally connected, it was originally thought that the leading and lagging strand polymerases would also be functionally coupled (Kornberg and Baker 1992). Specifically, if a lagging strand polymerase were to stall, the leading strand polymerase would also stop (and the converse). Indeed, early in vitro studies of the T4 phage and T7 phage replisomes indicated functional coupling of the two DNA polymerases (Lee et al. 1998; Salinas and Benkovic 2000). But studies in E. coli using both ensemble and single-molecule methods showed that the leading/lagging strand polymerases within the replisome are not functionally coupled, even though they are structurally attached (McInerney and O'Donnell 2004; Yao et al. 2009). A recent study has re-confirmed these actions in the E. coli system (Graham et al. 2017).

Eukaryotic Replisome

The CMG helicase tracks 3'–5' and thus encircles the leading strand, the opposite strand that bacterial DnaB helicase encircles (Ilves et al. 2010; Moyer et al. 2006). Furthermore, the leading strand polymerase epsilon directly binds to the replicative CMG helicase (Fig. 5.4) (Langston et al. 2014). This stands in contrast to the E. coli DNA polymerases which lack direct connection to the DnaB helicase, and require the clamp loader to mediate their connection. Also, unlike E. coli, the eukaryotic DNA polymerase alpha-primase forms an indirect contact to CMG mediated by the Ctf4 trimer replisome accessory factor (Simon et al. 2014; Sun et al. 2015). Furthermore, recent single-molecule studies demonstrate that DNA polymerase delta travels with the replisome for several kilobases (Lewis et al. 2020). These observations infer that numerous lagging strand DNA loops occur during replication of multiple Okazaki fragments in one DNA polymerase delta binding event, especially considering the small 100–200 bp length of eukaryotic Okazaki fragments (Smith and Whitehouse 2012). A strong interaction between the lagging strand DNA polymerase delta and individual replisome factors is not yet observed, and it seems possible that the connection of polymerase delta to the replisome may require multiple weak contacts that sum to form a strong connection. A previously identified contact between the Pol 32 subunit of DNA polymerase delta and the Pol 1 subunit of DNA polymerase alpha-primase (Huang et al. 1999; Johansson et al. 2004) has been shown to be functionally relevant in the single-molecule studies (Lewis et al. 2020).

Decades of textbook illustrations of the replisome show the helicase at the prow of the replication fork and the DNA polymerases trailing behind. However, structural

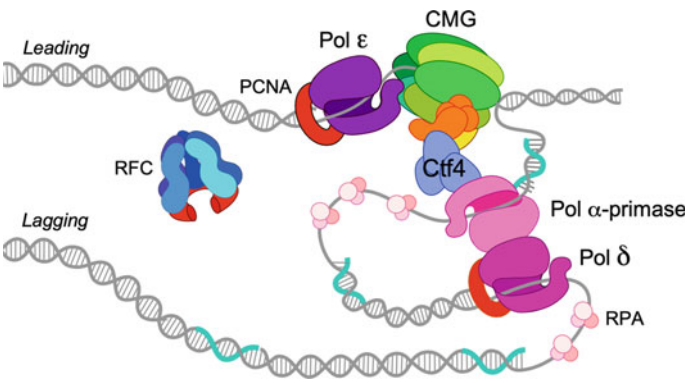


Fig. 5.4 Organization of the core replisome factors of the S. cerevisiae replisome. The leading strand polymerase epsilon is held to DNA by the PCNA clamp and by interaction with the CMG helicase that encircles the leading strand. The Ctf4 scaffolding trimer holds the lagging strand DNA polymerase alpha-primase to CMG helicase. The lagging strand DNA polymerase delta-PCNA complex travels with the replisome, resulting in 100-200 nucleotide Okazaki fragment DNA loops. The RPA heterotrimer binds the lagging strand ssDNA, protecting it from nucleases and melting secondary structures that impede polymerase progression

studies of the eukaryotic replisome reveal the leading strand DNA polymerase epsilon is on the opposite side of CMG from the lagging strand DNA polymerase alpha-primase, revealing that one polymerase travels ahead of the helicase and one travels behind the helicase (Fig. 5.4) (Sun et al. 2015). Further study revealed the orientation of CMG on DNA, and that polymerase alpha primase travels ahead of CMG while polymerase epsilon tracks in back of CMG (Georgescu et al. 2017). This anti-intuitive view of the lagging and leading strand DNA polymerases acting either in front or behind the helicase generalizes to the structure of the T7 bacteriophage replisome (Gao et al. 2019; Li and O'Donnell 2019). While the T7 bacteriophage does not use a clamp/clamp loader, it utilizes a hexameric helicase, and the structural studies reveal the identical leading and lagging strand polymerases are located on opposite faces of the helicase.

Now that the eukaryotic replisome structure is coming into view, one can compare and contrast several salient details of replisome architecture between E. coli and S. cerevisiae (compare Figs. 5.3 and 5.4). The directionality of the helicases places them on opposite strands, DnaB acts 5'–3'on the lagging strand and eukaryotic CMG acts 3'–5' on the leading strand. The CMG helicase organizes the eukaryotic replisome, binding the leading polymerase epsilon directly and the lagging DNA polymerase alpha-primase through the Ctf4 scaffolding factor. In E. coli, the leading and lagging strand DNA polymerases connect indirectly to the helicase through their mutual binding to the clamp loader that organizes the replisome. Bacterial primase does not form a stable attachment to the replisome, while eukaryotic DNA polymerase alpha-primase is attached to the replisome through the Ctf4 scaffolding factor (Gambus et al. 2009; Yuan et al. 2019). While eukaryotes utilize two distinct DNA polymerases for the two daughter strands, E. coli uses identical DNA polymerases. The bacterial primase synthesizes a RNA primer, while the eukaryotic DNA polymerase alpha-primase synthesizes a hybrid RNA-DNA primer. The eukaryotic CMG helicase is sandwiched, ahead and behind, by leading and lagging strand DNA polymerases, and DNA polymerases sandwiching the helicase generalizes to T7 bacteriophage, but whether this generalizes to bacterial cellular replisome is not yet certain.

Despite these differences, there are numerous similarities. The helicase motors in bacteria and eukaryotes are circular hexamers. The DNA polymerases in both cell types are tethered to DNA by circular clamps, loaded on DNA by a clamp loader pentamer. Furthermore, numerous Okazaki fragments can be synthesized by a single replisome in both bacteria and eukaryotes, indicating that the lagging strand polymerase stays attached to the replisome during multiple cycles of the lagging strand polymerase coming on/off DNA to extend Okazaki fragments in the opposite direction of fork movement. This "recycling" of the lagging strand polymerase is a fundamental problem for replisome action and is the subject of the next section.

Clamps Mediate Polymerase Recycling During Okazaki Fragment Synthesis

The antiparallel architecture of dsDNA, and unidirectional action of DNA polymerases require the lagging strand DNA polymerase to extend DNA in the opposite direction of the leading strand DNA polymerase. This action requires repeated priming and synthesis of the lagging strand in sections (Okazaki fragments) (Kornberg and Baker 1992). In vitro studies have elucidated an unanticipated mechanism by which this process is performed, as illustrated in Fig. 5.5.

Bacteria

In E. coli the repeated extension of RNA primers is accomplished by a DNA polymerase recycling process in which the same DNA polymerase is re-utilized numerous times by hopping among beta clamps, illustrated in Fig. 5.5 (Johnson and O'Donnell 2005). Specifically, while DNA polymerase III-beta is extending an Okazaki fragment, a new beta clamp is assembled on the next RNA primer by the clamp loader within the replisome. DNA polymerase III contains an inherent mechanism whereby it binds its beta clamp tightly during synthesis, but rapidly disengages from its beta

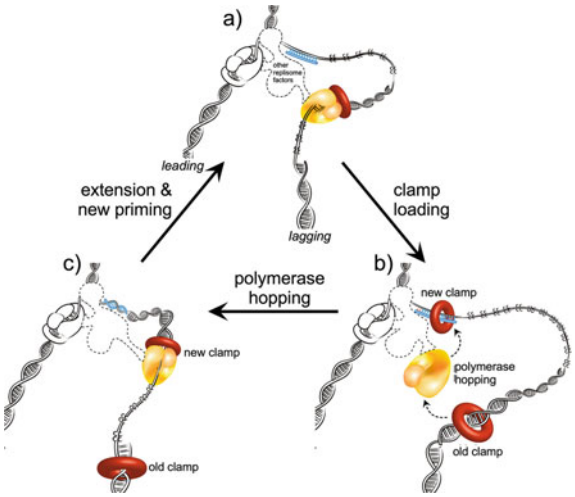


Fig. 5.5 Lagging strand polymerase recycling by hopping among sliding clamps. **a** The lagging strand DNA polymerase is attached to the leading strand DNA polymerase by connections with other replisome factors. The lagging strand is primed while the lagging strand polymerase extends a previously formed primer. **b** and **c** When the Okazaki fragment is complete, the lagging strand polymerase hops to a new clamp placed on the next primed site, leaving the old clamp behind on DNA

clamp upon finishing a section of DNA (O'Donnell 1987). Thus when polymerase III releases its beta clamp upon completing extension of an Okazaki fragment, it rapidly comes off DNA and binds a new beta clamp at the next primed site for rapid and processive extension of the next Okazaki fragment (Stukenberg et al. 1994). This rapid release and recycling process is referred to as "collision release".

Further work in the T4 phage and E. coli systems revealed that not all lagging strand DNA polymerase recycling events require the complete extension of an Okazaki fragment, and that sometimes the DNA polymerase is "signaled" to recycle from one clamp to a new clamp before an Okazaki fragment is extended to completion (Kurth et al. 2013; Li and Marians 2000; Yang et al. 2004). While the signal of this "signal release" mechanism is not clear, the process of disengaging the strong DNA polymerase-clamp connection while extending DNA would require significant energy. Torsional strain during replication contains more than sufficient energy to disrupt the strong polymerase-clamp connection, and experimental evidence exists that the extensive energy generated by supercoiling during fork advance can force the dissociation of polymerase from its clamp before an Okazaki fragment is complete (Kurth et al. 2013).

Single-molecule assays have shown that recycling to new clamps by a lagging strand polymerase that remains attached to the replisome can operate for at least 86 kb of fork progression, during which multiple Okazaki fragments of 1 kb or less are generated (Yao et al. 2009). Polymerase hopping among sliding clamps (O'Donnell 1987; Stukenberg et al. 1994) is the accepted process by which E. coli DNA polymerase III recycles from one Okazaki fragment to the next. This process appears to be at work in eukaryotes as well, and is described in more detail in the next section.

Eukaryotes

The lagging strand specific eukaryotic DNA polymerase delta-PCNA clamp is highly processive during replication of a > 5 kb primed substrate, and despite the non-homology in bacterial and eukaryotic DNA polymerases, DNA polymerase delta demonstrates collision release from its PCNA clamp upon completing DNA much as observed for E. coli DNA polymerase III and illustrated in Fig. 5.5 (Langston and O'Donnell 2008). However, unlike the E. coli system a strong and direct connection of polymerase delta to other replisome components has not been detected, and therefore it is generally thought that the eukaryotic lagging strand polymerase delta is used stoichiometrically—one DNA polymerase delta molecule for each Okazaki fragment. As described earlier in this review, a recent single-molecule study of the eukaryotic replisome reveals that a single molecule of DNA polymerase delta remains with the replisome for numerous Okazaki fragments, and therefore is efficiently recycled during replication (Lewis et al. 2020). Interestingly, when additional DNA polymerase delta is present in solution there occurs an exchange of polymerase delta from

solution with polymerase delta in the replisome. DNA polymerases epsilon and alpha-primase also undergo a similar exchange rate when either of these DNA polymerases is present in excess (Lewis et al. 2020). This result is consistent with observations of protein exchange within bacterial replisomes that depends on excess unbound protein (Mueller et al. 2019). Regardless, the single molecule studies utilizing an 18 kb DNA in the absence of excess solution phase proteins reveals that polymerase delta can be held into the replisome for numerous Okazaki fragments that requires DNA looping as illustrated in Fig. 5.5, especially considering that each eukaryotic Okazaki fragment is only 100–200 bp in length (Lewis et al. 2020; Smith and Whitehouse 2012). One connection point of DNA polymerase delta is through DNA polymerase alpha-primase (Huang et al. 1999), consistent with polymerase delta function in extending lagging strand primed sites into full length Okazaki fragments (Lewis et al. 2020).

It is unclear why DNA looping is utilized in DNA replication, either bacterial or eukaryotic. Perhaps DNA looping on the lagging strand is simply a consequence of a processive lagging strand polymerase that is held to the replication fork. Whether there are other advantages to a DNA looping process for lagging strand synthesis will remain for future studies.

Assembly and Escape of Replicative Helicases from the Origin

The goal of an origin, in both bacteria and eukaryotes, is to load two helicases onto the DNA to form bidirectional replication forks. In bacteria this is a comparatively simple process because they often have one circular chromosome with only one origin, and the firing of the origin is not strictly regulated. For example, in rich growth media the single origin in the E. coli chromosome can fire more than one time before cell division takes place (Kornberg and Baker 1992). Unlike E. coli, eukaryotes contain numerous origins within each of several linear chromosomes, and unregulated firing of these numerous origins could lead to genome instability. In fact, eukaryotic origins are highly regulated, being "licensed" in the G1 phase of the cell cycle, constraining them to fire once and only once during the ensuing S phase of the cell cycle and preventing re-initiations that could result in a tangle of DNA segments having different copy numbers (Blow and Laskey 1988). Many studies on this subject have led to the finding of numerous proteins and two cell cycle kinases that are required for initiation at eukaryotic origins (Bell and Labib 2016). Study of the licensing step in G1 phase revealed that two Mcm2–7 hexamers are loaded onto dsDNA at S. cerevisiae origins (Evrin et al. 2009; Remus et al. 2009). Upon entering S-phase, origin initiation factors assemble GINS and Cdc45 onto each Mcm2–7 to form two CMG helicases. This two-step process, separated into two cell cycle stages, ensures than an origin only fires once per cell cell cycle.

Bacterial Helicase

The objective of the bacterial origin is to load two helicases onto DNA. However, while eukaryotic CMG is assembled around dsDNA and must melt dsDNA to switch onto ssDNA, the bacterial helicase is directly loaded onto a ssDNA bubble formed by origin binding proteins (e.g. DnaA protein) (Bramhill and Kornberg 1988a, b). Also, unlike eukaryotic CMG, the bacterial replicative hexameric helicase tracks C-tier first (Itsathitphaisarn et al. 2012; Lyubimov et al. 2011). In many bacteria the helicase, like E. coli DnaB helicase, exists as a hexameric ring that use a protein (e.g. E. coli DnaC) that acts as a "ring breaker" to open and close the DnaB ring around the ssDNA, but some bacteria have proteins that act as "ring makers" that assemble the hexameric helicase around ssDNA from monomeric subunits (Davey and O'Donnell 2003). In either case, once the helicase ring is assembled and encircles ssDNA it can proceed to unwind dsDNA by the steric exclusion process. It is interesting to note that the two E. coli DnaB helicases assembled at the open ssDNA bubble of an origin face one another and must pass each another during initial origin unwinding (Fang et al. 1999). As described below, this aspect mimics the two eukaryotic CMG helicases that pass one another in the final stages of bidirectional origin unwinding.

Eukaryotic Helicase

Origins in eukaryotes require numerous proteins that act as chaperonin assembly factors to assemble the 11 subunit CMG helicase at the G1/S phase boundary. At this boundary, there exist numerous factors and kinases that modify and assemble head-to-head CMGs to ensure that an origin fires once and only once per cell cycle. Most of these factors are needed to assemble the 11-subunit CMG helicase, but do not travel with the replication fork (Bell and Labib 2016). It has long been thought that eukaryotic CMG tracks C-tier first like the bacterial hexameric helicase (Bell and Labib 2016). This view is promoted by the fact that the two CMGs that are formed at an origin are oriented N-tier to N-tier, suggesting they could simply move apart in opposite directions to form bidirectional replication forks if they traveled C-tier first.

Despite the attractive view of CMGs moving C-tier first from an origin, a C-tier first orientation of CMG unwinding, when modeled into the cryo-EM replisome structure, would require the leading strand polymerase epsilon to ride at the top of the replisome during replication, and thus the unwound DNA would need to loop back from the bottom of CMG to the top of the replisome to reach polymerase epsilon, a distance of approximately 120 Angstroms and would leave the ssDNA vulnerable to nuclease attack (Sun et al. 2015). Use of a strategy to determine a high resolution cryo-EM 3D structure of CMG acting at a replication fork revealed that CMG tracks N-tier first along ssDNA, which contradicted the long-held view that CMG travels C-tier first (Georgescu et al. 2017). This new "N-tier first orientation of CMG" showed

that the two CMGs at an origin are directed inward toward one another, not outward. While inward directed CMGs would block one another, and may initially seem a stand-off, this view of an inward direction of CMGs also offers a new way to look at origin initiation (Georgescu et al. 2017). Specifically, if each CMG binds one strand while surrounding dsDNA, the two opposite facing CMGs will block one another from moving, but would pull on opposite strands. If they have sufficient force, they could rip the DNA strands apart by pulling them through their respective rings while the two CMG rings stand still. In fact, CMG has been demonstrated to track on only one strand while encircling the duplex (Langston and O'Donnell 2019). Interestingly, earlier studies of E. coli DnaB and eukaryotic Mcm4,6,7 showed that they could rip DNA duplex structures apart, but the exact function of this observation was not understood at the time (Kaplan et al. 2003; Kaplan and O'Donnell 2002). Recent work using CMG and a 150 bp linear ARS1 origin DNA showed that two head-on CMGs can in fact rip, or shear, the DNA strands apart provided that Mcm10 is also present (Langston and O'Donnell 2019). Mcm10 binds tightly to CMG and enhances helicase activity over 10-fold (Langston et al. 2017). Furthermore, in vivo studies document that Mcm10 is required for head-on CMGs to unwind the origin (Kanke et al. 2012; van Deursen et al. 2012; Watase et al. 2012). Hence, two head-on CMGs that track on opposite strands may explain how head-to-head CMGs + Mcm10 unwind dsDNA at origins (Fig. 5.6).

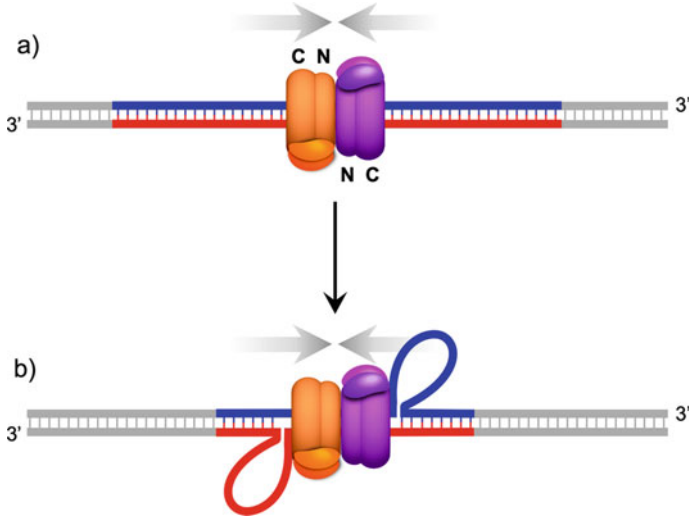


Fig. 5.6 Head-to-head CMGs shear dsDNA to ssDNA. Two CMGs are oriented head-to-head (N-tier to N-tier) on dsDNA at eukaryotic origins. The CMGs track N-first on DNA and thus the two CMGs are directed inward toward one another. Furthermore, each CMG tracks on one one strand of the duplex and therefore pull the two strands in opposite directions. Experiments show that the pulling force is sufficient to shear apart the two strands of a 150 bp ARS1 origin of yeast (Langston and O'Donnell 2019)

Once CMGs that encircle dsDNA unwind an origin, they must transition to ssDNA to become bone fide replicative helicases acting at replication forks (Fig. 5.7). The mechanism of this piece of the initiation puzzle has been determined from a single-molecule study (Wasserman et al. 2019). In the single-molecule study, CMG + Mcm10 could load onto linear ssDNA that is blocked by optical traps on either end. This indicates that CMG contains a ssDNA gate to enable it to get onto ssDNA, because both ends of the DNA are blocked by beads in the optical trap. The inference of a ssDNA gate in CMG is also consistent with earlier studies demonstrating that CMG can melt a 5' tailed oligonucleotide from circular M13 ssDNA (Ilves et al. 2010). Cryo-EM structures of CMG show no openings in the Mcm2-7 ring (Abid Ali et al. 2016; Yuan et al. 2016) and therefore one may presume the ssDNA gate is transient. The presence of a ssDNA gate in CMG also has implications to DNA repair, to which the reader is referred to the original article (Wasserman et al. 2019).

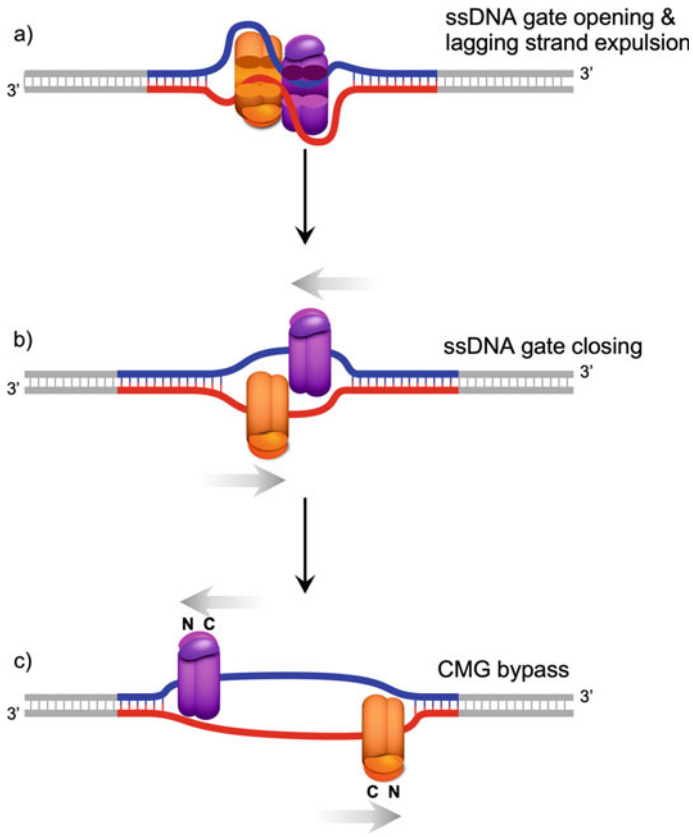


Fig. 5.7 A ssDNA gate in CMG enables it to transition from dsDNA to ssDNA. CMG is assembled onto dsDNA at origins, but must transition onto ssDNA for steric exclusion helicase action at replication forks. CMG is demonstrated to have a ssDNA gate that allows it to exclude one strand after initial unwinding of dsDNA. See text for details

The CMG does not load onto dsDNA, consistent with the Mcm2–5 dsDNA entry gate of Mcm2–7 being shut tight by the Cdc45/GINS accessory factors (Abid Ali et al. 2016; Yuan et al. 2016). Hence, the ssDNA gate is presumed to be located between Mcm subunits other than the Mcm2/5 interface, and identification of this ssDNA gate awaits future studies.

Assembly of Bacterial and Eukaryotic Replisomes

Bacterial Replisome

Two DnaB helicases assemble onto the ssDNA bubble formed by origin binding factors (E. coli DnaA/DnaC), and a further small amount of unwinding (< 100 bp) by the DnaB helicases enables primase to form RNA primed sites (Bramhill and Kornberg 1988b; Fang et al. 1999). The DnaB helicase also recruits the clamp loader through direct interaction (Kim et al. 1996). These actions only require the two bidirectional helicases to unwind less than 100 bp at an origin to enable priming of both strands and to assemble two replisomes (Fang et al. 1999). Replisome assembly is straightforward because the E. coli clamp loader tightly binds three DNA polymerase III molecules to form a tightly associated assembly containing one clamp loader and three DNA polymerase III's (McInerney et al. 2007). Thus, once DnaB recruits a primase to form a primed site, the clamp loader within the clamp loader-3 polymerase III assembly only needs to place clamps onto the primed DNA and bind the helicase and clamp for assembly of a leading/lagging strand replisome. After assembly of the helicase onto ssDNA, the replisome assembly process occurs spontaneously and without need of catalytic assembly factors (McInerney et al. 2007).

Eukaryotic Replisome

The fact that eukaryotes have different DNA polymerases for the leading and lagging strands might suggest the need for certain chaperonin factors and/or kinases to correctly assemble it, much like the assembly of CMG helicase at an origin. The process of asymmetric assembly of the replicative polymerases has been recapitulated in vitro (Georgescu et al. 2014; Georgescu et al. 2015b). The process involves two suppression reactions that eliminate a DNA polymerase from the replication fork if they load onto the wrong strand. The asymmetric polymerase assembly process also involves two major recruitment interactions that are specific for each DNA polymerase. The recruitment mechanism that locates DNA polymerase delta to the lagging strand is due to its much stronger affinity to the PCNA clamp compared to the affinity of DNA polymerase epsilon to PCNA (Schauer et al. 2017). The suppression mechanism that prevents polymerase epsilon from function on the lagging strand is

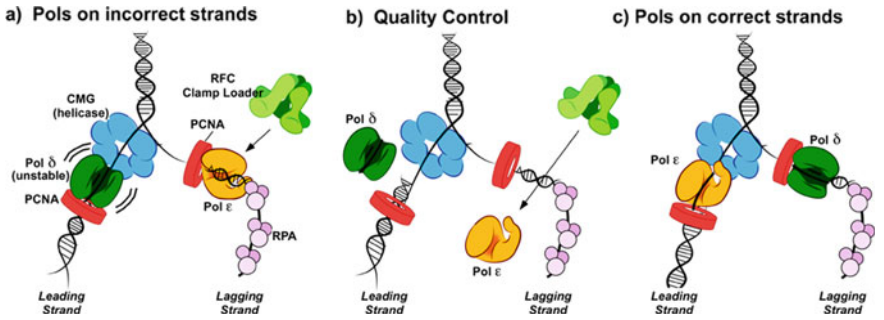


Fig. 5.8 Quality control processes that direct DNA polymerases epsilon and delta to their respective strands. Panels **a** and **b** When a DNA polymerase loads onto the wrong strand, it is unstable. Thus DNA polymerase delta undergoes collision release from PCNA-DNA upon colliding with CMG on the leading strand. The RFC clamp loader out-competes polymerase epsilon on lagging strand primed sites. **c** When a polymerase assembles on the correct strand, it is stabilized and retained. Thus polymerase epsilon binds to CMG and is held to the leading strand, and CMG also prevents RFC from displacing polymerase epsilon from a leading strand primed site. Polymerase delta binds tightly to PCNA at a primed site on the lagging strand, preventing it from becoming displaced by the RFC clamp loader. Reproduced with permission from Fig. 5.8 in (Schauer et al. 2017)

its sensitivity to inhibition by the RFC clamp loader, which competes DNA polymerase epsilon off a primer terminus (Schauer et al. 2017). This RFC suppression does not occur on the leading strand because CMG binds DNA polymerase epsilon and protects it from RFC inhibition, thus enabling PCNA to be loaded onto the leading strand for use by DNA polymerase epsilon (Schauer et al. 2017). The mechanism that locates DNA polymerase epsilon to the leading strand is also two-fold. Firstly, DNA polymerase epsilon is recruited to the leading strand by specific interaction with CMG that encircles leading strand ssDNA (Langston et al. 2014). Once polymerase epsilon is complexed to CMG, polymerase delta cannot displace it from the leading strand (Georgescu et al. 2014; Georgescu et al. 2015b). But if DNA polymerase delta arrives at the leading strand before DNA polymerase epsilon, DNA polymerase delta self-ejects from DNA and PCNA upon colliding with CMG (Georgescu et al. 2014; Schauer et al. 2017). Taken together the studies indicate that assembly of an asymmetric eukaryotic replisome is inherent in the core factors and does not require outside assembly factors as illustrated in Fig. 5.8. Suppression reactions for polymerase occupancy are unprecedented for bacterial systems, because the leading and lagging strand DNA polymerases are identical.

Summary and Future Outlook

Reconstitution of the bacterial and eukaryotic replisomes reveal the main structural and biochemical differences by which the replisome machine functions in these two cell types. The non-homologous sequences and structures of the helicase, primase

and DNA polymerase core replisome factors of bacteria and eukaryotes is compelling evidence that they evolved separately. However, as explained here, there are uncanny similarities between assembly and function of bacterial and eukaryotic replisomes when viewed with a wide-angle lens. Specifically, the origins have the same goal of loading two bidirectional helicases onto DNA. Both the bacterial and eukaryotic helicases are hexamers with N- and C-tiers. But bacterial and eukaryotic hexameric helicases act on opposite strands setting the ground work for a distinct organization of their respective replisomes. In bacteria, the clamp loader binds the helicase and joins it to the leading and lagging strand DNA polymerases. In eukaryotes the helicase directly binds the leading strand DNA polymerase, and binds the lagging strand DNA polymerase indirectly through other replisome proteins. In both cases, two replisomes spontaneously assemble once the helicases are loaded onto a bidirectional origin. Another similarity is that in both cell types, the mechanism of polymerase recycling by hopping among DNA sliding clamps for synthesis of lagging strand Okazaki fragments appears to be conserved, although future studies are need to confirm if this is true.

These generalities are comforting for the idea that an understanding of replisome action is at hand, but there are still many mysteries that are yet to be solved. The reason that the eukaryotic replisome uses different DNA polymerases for leading and lagging strand DNA synthesis is unknown, while bacteria utilize identical enzymes for the same purpose. Also, yet to be determined is the reason that eukaryotes use a RNA-DNA primer to initiate DNA synthesis by the replicative DNA polymerases epsilon and delta. Could it be that the RNA portion is all that is normally used, as in bacteria and its phages? Perhaps the Pol 1 subunit activity of polymerase alpha-primase is only required for certain regions of the genome. Another remaining mystery is why the eukaryotic helicase has so many subunits. Presumably the extra subunits are needed after the initial licensing step to close the gap between Mcm2 and Mcm5, needed for assembly of the Mcm2–7 ring onto dsDNA in G1 phase. However, these accessory subunits also bind DNA polymerase epsilon and the Ctf4 scaffolding factor, and thus appear central to direct the leading and lagging strand DNA polymerases to their respective strands. It also remains possible that the multisubunit structure of CMG is used for regulation, such as in response to the S phase replication checkpoint which involves modification of some of the CMG subunits (Ilves et al. 2012; Labib and De Piccoli 2011).

Eukaryotic replisomes must also deal with nucleosomes, epigenetic marks, sister chromosome cohesion, and DNA repair. These aspects have not been dealt with in this review. While it is obvious that replication must interface with these important processes, there is still very little known about the mechanistic details. For example, it has been shown in budding yeast that mutations in histone binding sites in DNA polymerase epsilon, Pol 1 of DNA polymerase alpha-primase, and an N-terminal region of Mcm2 result in differential transfer of parental nucleosomes to either the leading or lagging daughter strand, depending on the mutant protein (Gan et al. 2018; He et al. 2017; Petryk et al. 2018; Yu et al. 2018). These findings imply there are different pathways for epigenetic marks to be transferred to one strand or the other during genome replication. The mechanistic basis for these pathways

are unknown, and remain important open questions for future studies with high potential for an impact on disease and on developmental biology. While there is very little is known about how replication interfaces with these important processes, a mechanistic understanding of these aspects of chromatin structure and replication are very important to our understanding of development, cancer, and ultimately a cure for certain human health and diseases.

References

Abid Ali F, Renault L, Gannon J, Gahlon HL, Kotecha A, Zhou JC, Rueda D, Costa A (2016) Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. Nat Commun 7:10708. https://doi.org/10.1038/ncomms10708

Ahnert P, Patel SS (1997) Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. J Biol Chem 272(51):32267–32273

Antony E, Lohman TM (2019) Dynamics of E. coli single stranded DNA binding (SSB) protein-DNA complexes. Semin Cell Dev Biol 86:102–111. https://doi.org/10.1016/j.semcdb.2018.03.017

Aravind L, Leipe DD, Koonin EV (1998) Toprim–a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res 26(18):4205–4213. https://doi.org/10.1093/nar/26.18.4205

Bell SP, Labib K (2016) Chromosome duplication in saccharomyces cerevisiae. Genetics 203(3):1027–1067. https://doi.org/10.1534/genetics.115.186452

Blow JJ, Laskey RA (1988) A role for the nuclear envelope in controlling DNA replication within the cell cycle. Nature 332(6164):546–548. https://doi.org/10.1038/332546a0

Bramhill D, Kornberg A (1988a) Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell 52(5):743–755

Bramhill D, Kornberg A (1988b) A model for initiation at origins of DNA replication. Cell 54(7):915–918. https://doi.org/10.1016/0092-8674(88)90102-x

Burgers PMJ, Kunkel TA (2017) Eukaryotic DNA Replication Fork. Annu Rev Biochem 86:417–438. https://doi.org/10.1146/annurev-biochem-061516-044709

Chen R, Wold MS (2014) Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. Bioessays 36(12):1156–1161. https://doi.org/10.1002/bies.201400107

Conaway RC, Lehman IR (1982) A DNA primase activity associated with DNA polymerase alpha from Drosophila melanogaster embryos. Proc Natl Acad Sci U S A 79(8):2523–2527

Cullmann G, Fien K, Kobayashi R, Stillman B (1995) Characterization of the five replication factor C genes of Saccharomyces cerevisiae. Mol Cell Biol 15(9):4661–4671

Dallmann HG, Kim S, Pritchard AE, Marians KJ, McHenry CS (2000) Characterization of the unique C terminus of the Escherichia coli tau DnaX protein. Monomeric C-tau binds alpha AND DnaB and can partially replace tau in reconstituted replication forks. J Biol Chem 275(20):15512–15519. https://doi.org/10.1074/jbc.M909257199 M909257199 [pii]

Davey MJ, O'Donnell M (2003) Replicative helicase loaders: ring breakers and ring makers. Curr Biol 13(15):R594–596. https://doi.org/10.1016/s0960-9822(03)00523-2

De Biasio A, Blanco FJ (2013) Proliferating cell nuclear antigen structure and interactions: too many partners for one dancer? Adv Protein Chem Struct Biol 91:1–36. https://doi:B978-0-12-411637-5.00001-9 [pii] https://doi.org/10.1016/b978-0-12-411637-5.00001-9

Dua R, Levy DL, Campbell JL (1999) Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. J Biol Chem 274(32):22283–22288

Enemark EJ, Joshua-Tor L (2008) On helicases and other motor proteins. Curr Opin Struct Biol 18(2):243–257. https://doi:S0959-440X(08)00011-0 [pii] https://doi.org/10.1016/j.sbi.2008.01.007

Evrin C, Clarke P, Zech J, Lurz R, Sun J, Uhle S, Li H, Stillman B, Speck C (2009) A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc Natl Acad Sci U S A 106(48):20240–20245. https://doi:0911500106 [pii] https://doi.org/10.1073/pnas.0911500106

Fang L, Davey MJ, O'Donnell M (1999) Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin. Mol Cell 4(4):541–553

Forterre P, Filee J, Myllykallio H (2004) Origin and evolution of DNA and DNA replication machineries. In: de Pouplana LR (ed) The genetic code and the origin of life. Landes Bioscience, pp. 145–168

Fu YV, Yardimci H, Long DT, Ho TV, Guainazzi A, Bermudez VP, Hurwitz J, van Oijen A, Scharer OD, Walter JC (2011) Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. Cell 146(6):931–941. https://doi:S0092-8674(11)00889-0 [pii] https://doi.org/10.1016/j.cell.2011.07.045

Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol 8(4):358–366. https://doi.org/10.1038/ncb1382

Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, Calzada A, Labib K (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. EMBO J 28(19):2992–3004. https://doi.org/10.1038/emboj.2009.226

Gan H, Serra-Cardona A, Hua X, Zhou H, Labib K, Yu C, Zhang Z (2018) The Mcm2-Ctf4-Polalpha Axis Facilitates Parental Histone H3-H4 Transfer to Lagging Strands. Mol Cell 72(1):140–151 e143. https://doi.org/10.1016/j.molcel.2018.09.001

Gao D, McHenry CS (2001) tau binds and organizes Escherichia coli replication through distinct domains. Partial proteolysis of terminally tagged tau to determine candidate domains and to assign domain V as the alpha binding domain. J Biol Chem 276(6):4433–4440. https://doi.org/10.1074/jbc.M009828200

Gao Y, Cui Y, Fox T, Lin S, Wang H, de Val N, Zhou ZH, Yang W (2019) Structures and operating principles of the replisome. Science 363(6429). https://doi.org/10.1126/science.aav7003

Georgescu RE, Langston L, Yao NY, Yurieva O, Zhang D, Finkelstein J, Agarwal T, O'Donnell ME (2014) Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. Nat Struct Mol Biol 21(8):664–670. https://doi:nsmb.2851 [pii] https://doi.org/10.1038/nsmb.2851

Georgescu R, Langston L, O'Donnell M (2015a) A proposal: Evolution of PCNA's role as a marker of newly replicated DNA. DNA Repair (Amst) 29:4–15. https://doi:S1568-7864(15)00027-0 [pii] https://doi.org/10.1016/j.dnarep.2015.01.015

Georgescu RE, Schauer GD, Yao NY, Langston LD, Yurieva O, Zhang D, Finkelstein J, O'Donnell ME (2015b) Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. Elife 4:e04988. https://doi.org/10.7554/eLife.04988

Georgescu R, Yuan Z, Bai L, de Luna Almeida Santos R, Sun J, Zhang D, Yurieva O, Li H, O'Donnell ME (2017) Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. Proc Natl Acad Sci U S A 114(5):E697–E706. https://doi.org/10.1073/pnas.1620500114

Goswami P, Abid Ali F, Douglas ME, Locke J, Purkiss A, Janska A, Eickhoff P, Early A, Nans A, Cheung AMC, Diffley JFX, Costa A (2018) Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. Nat Commun 9(1):5061. https://doi.org/10.1038/s41467-018-07417-1

Graham JE, Marians KJ, Kowalczykowski SC (2017) Independent and Stochastic Action of DNA Polymerases in the Replisome. Cell 169(7):1201–1213 e1217. https://doi.org/10.1016/j.cell.2017.05.041

Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1996) Structure of the C-terminal region of p 21(WAF1/CIP1) complexed with human PCNA. Cell 87(2):297–306. https://doi.org/10.1016/S0092-8674(00)81347-1

Hacker KJ, Johnson KA (1997) A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding. Biochemistry 36(46):14080–14087. https://doi.org/10.1021/bi971644v bi971644v [pii]

He H, Li Y, Dong Q, Chang AY, Gao F, Chi Z, Su M, Zhang F, Ban H, Martienssen R, Chen YH, Li F (2017) Coordinated regulation of heterochromatin inheritance by Dpb3-Dpb4 complex. Proc Natl Acad Sci U S A 114(47):12524–12529. https://doi.org/10.1073/pnas.1712961114

Hogg M, Osterman P, Bylund GO, Ganai RA, Lundstrom EB, Sauer-Eriksson AE, Johansson E (2014) Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. Nat Struct Mol Biol 21(1):49–55. https://doi.org/10.1038/nsmb.2712

Huang ME, Le Douarin B, Henry C, Galibert F (1999) The Saccharomyces cerevisiae protein YJR043C (Pol32) interacts with the catalytic subunit of DNA polymerase alpha and is required for cell cycle progression in G2/M. Mol Gen Genet 260(6):541–550. https://doi.org/10.1007/s004380050927

Ilves I, Petojevic T, Pesavento JJ, Botchan MR (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell 37(2):247–258. https://doi.org/10.1016/j.molcel.2009.12.030

Ilves I, Tamberg N, Botchan MR (2012) Checkpoint kinase 2 (Chk2) inhibits the activity of the Cdc45/MCM2-7/GINS (CMG) replicative helicase complex. Proc Natl Acad Sci USA 109(33):13163–13170. https://doi:1211525109 [pii] https://doi.org/10.1073/pnas.1211525109

Itsathitphaisarn O, Wing RA, Eliason WK, Wang J, Steitz TA (2012) The hexameric helicase DnaB adopts a nonplanar conformation during translocation. Cell 151(2):267–277. https://doi:S0092-8674(12)01117-8 [pii] https://doi.org/10.1016/j.cell.2012.09.014

Jeruzalmi D, O'Donnell M, Kuriyan J (2001) Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. Cell 106(4):429-441. https://doi:S0092-8674(01)00463-9 [pii]

Johansson E, Garg P, Burgers PM (2004) The Pol32 subunit of DNA polymerase delta contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. J Biol Chem 279(3):1907–1915. https://doi.org/10.1074/jbc.M310362200 M310362200 [pii]

Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 74:283–315. https://doi.org/10.1146/annurev.biochem.73.011303.073859

Kanke M, Kodama Y, Takahashi TS, Nakagawa T, Masukata H (2012) Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. EMBO J 31 (9):2182–2194. https://doi:emboj201268 [pii] https://doi.org/10.1038/emboj.2012.68

Kaplan DL, O'Donnell M (2002) DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands. Mol Cell 10(3):647–657

Kaplan DL, Davey MJ, O'Donnell M (2003) Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. J Biol Chem 278(49):49171–49182. https://doi.org/10.1074/jbc.M308074200 M308074200 [pii]

Kelch BA (2016) Review: the lord of the rings: Structure and mechanism of the sliding clamp loader. Biopolymers 105(8):532–546. https://doi.org/10.1002/bip.22827

Kelch BA, Makino DL, O'Donnell M, Kuriyan J (2011) How a DNA polymerase clamp loader opens a sliding clamp. Science 334(6063):1675–1680. https://doi.org/10.1126/science.1211884

Kelch BA, Makino DL, O'Donnell M, Kuriyan J (2012) Clamp loader ATPases and the evolution of DNA replication machinery. BMC Biol 10:34. https://doi.org/10.1186/1741-7007-10-34

Kesti T, Flick K, Keranen S, Syvaoja JE, Wittenberg C (1999) DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. Mol Cell 3(5):679–685. https://doi:S1097-2765(00)80361-5 [pii]

Kilkenny ML, Longo MA, Perera RL, Pellegrini L (2013) Structures of human primase reveal design of nucleotide elongation site and mode of Pol alpha tethering. Proc Natl Acad Sci USA 110(40):15961–15966. https://doi.org/10.1073/pnas.1311185110

Kim S, Dallmann HG, McHenry CS, Marians KJ (1996) Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. Cell 84(4):643–650. https://doi:S0092-8674(00)81039-9 [pii]

Kirk BW, Kuchta RD (1999) Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology. Biochemistry 38(24):7727–7736. https://doi.org/10.1021/bi990247c

Kong XP, Onrust R, O'Donnell M, Kuriyan J (1992) Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. Cell 69(3):425–437. https://doi.org/10.1016/0092-8674(92)90445-I

Koonin EV, Dolja VV, Krupovic M (2015) Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479–480:2–25. https://doi:S0042-6822(15)00085-9 [pii] https://doi.org/10.1016/j.virol.2015.02.039

Kornberg A, Baker TA (1992) DNA Replication. Freeman, New York, W.H

Kurth I, Georgescu RE, O'Donnell ME (2013) A solution to release twisted DNA during chromosome replication by coupled DNA polymerases. Nature 496(7443):119–122. https://doi:nature11988 [pii] https://doi.org/10.1038/nature11988

Labib K, De Piccoli G (2011) Surviving chromosome replication: the many roles of the S-phase checkpoint pathway. Philos Trans R Soc Lond B Biol Sci 366(1584):3554–3561. https://doi:366/1584/3554 [pii] https://doi.org/10.1098/rstb.2011.0071

Langston LD, O'Donnell M (2008) DNA polymerase delta is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. J Biol Chem 283(43):29522–29531. https://doi.org/10.1074/jbc.M804488200

Langston LD, O'Donnell ME (2019) An explanation for origin unwinding in eukaryotes. Elife 8. https://doi.org/10.7554/eLife.46515

Langston LD, Zhang D, Yurieva O, Georgescu RE, Finkelstein J, Yao NY, Indiani C, O'Donnell ME (2014) CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. Proc Natl Acad Sci U S A 111(43):15390–15395. https://doi:1418334111 [pii] https://doi.org/10.1073/pnas.1418334111

Langston LD, Mayle R, Schauer GD, Yurieva O, Zhang D, Yao NY, Georgescu RE, O'Donnell ME (2017) Mcm10 promotes rapid isomerization of CMG-DNA for replisome bypass of lagging strand DNA blocks. eLife 6:e29118. https://doi.org/10.7554/eLife.29118

LeBowitz JH, McMacken R (1986) The Escherichia coli dnaB replication protein is a DNA helicase. J Biol Chem 261(10):4738–4748

Lee J, Chastain PD, 2nd, Kusakabe T, Griffith JD, Richardson CC (1998) Coordinated leading and lagging strand DNA synthesis on a minicircular template. Mol Cell 1(7):1001–1010. https://doi:S1097-2765(00)80100-8 [pii]

Lee SJ, Syed S, Enemark EJ, Schuck S, Stenlund A, Ha T, Joshua-Tor L (2014) Dynamic look at DNA unwinding by a replicative helicase. Proc Natl Acad Sci USA 111(9):E827–835. https://doi:1322254111 [pii] https://doi.org/10.1073/pnas.1322254111

Lewis JS, Spenkelink LM, Schauer GD, Yurieva O, Mueller SH, Natarajan V, Kaur G, Maher C, Kay C, O'Donnell ME, van Oijen AM (2020) Tunability of DNA Polymerase Stability during Eukaryotic DNA Replication. Mol Cell 77(1):17–25 e15. https://doi.org/10.1016/j.molcel.2019.10.005

Li X, Marians KJ (2000) Two distinct triggers for cycling of the lagging strand polymerase at the replication fork. J Biol Chem 275(44):34757–34765. https://doi.org/10.1074/jbc.M006556200

Li H, O'Donnell ME (2019) DNA replication from two different worlds. Science 363(6429):814–815. https://doi.org/10.1126/science.aaw6265

Lyubimov AY, Strycharska M, Berger JM (2011) The nuts and bolts of ring-translocase structure and mechanism. Curr Opin Struct Biol 21(2):240–248. https://doi.org/10.1016/j.sbi.2011.01.002
Makarova KS, Koonin EV, Kelman Z (2012) The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. Biol Direct 7:7. https://doi:1745-6150-7-7 [pii] https://doi.org/10.1186/1745-6150-7-7
Makarova KS, Krupovic M, Koonin EV (2014) Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. Front Microbiol 5:354. https://doi.org/10.3389/fmicb.2014.00354
McHenry CS (2011) Bacterial replicases and related polymerases. Curr Opin Chem Biol 15(5):587–594. https://doi:S1367-5931(11)00125-6 [pii] https://doi.org/10.1016/j.cbpa.2011.07.018
McInerney P, O'Donnell M (2004) Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. J Biol Chem 279(20):21543–21551. https://doi.org/10.1074/jbc.M401649200 M401649200 [pii]
McInerney P, Johnson A, Katz F, O'Donnell M (2007) Characterization of a triple DNA polymerase replisome. Mol Cell 27(4):527–538. https://doi.org/10.1016/j.molcel.2007.06.019
Moyer SE, Lewis PW, Botchan MR (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc Natl Acad Sci USA 103(27):10236–10241. https://doi.org/10.1073/pnas.0602400103
Mueller SH, Spenkelink LM, van Oijen AM (2019) When proteins play tag: the dynamic nature of the replisome. Biophys Rev 641–651. https://doi.org/10.1007/s12551-019-00569-4
O'Donnell ME (1987) Accessory proteins bind a primed template and mediate rapid cycling of DNA polymerase III holoenzyme from Escherichia coli. J Biol Chem 262(34):16558–16565
O'Donnell M, Jeruzalmi D, Kuriyan J (2001) Clamp loader structure predicts the architecture of DNA polymerase III holoenzyme and RFC. Curr Biol 11(22):R935–946. https://doi:S0960-9822(01)00559-0 [pii]
O'Donnell ME, Li H (2018) The ring-shaped hexameric helicases that function at DNA replication forks. Nat Struct Mol Biol 25(2):122–130. https://doi.org/10.1038/s41594-018-0024-x
O'Donnell M, Onrust R, Dean FB, Chen M, Hurwitz J (1993) Homology in accessory proteins of replicative polymerases–E. coli to humans. Nucleic Acids Res 21(1):1–3. https://doi.org/10.1093/nar/21.1.1
Petryk N, Dalby M, Wenger A, Stromme CB, Strandsby A, Andersson R, Groth A (2018) MCM2 promotes symmetric inheritance of modified histones during DNA replication. Science 361(6409):1389–1392. https://doi.org/10.1126/science.aau0294
Podobnik M, McInerney P, O'Donnell M, Kuriyan J (2000) A TOPRIM domain in the crystal structure of the catalytic core of Escherichia coli primase confirms a structural link to DNA topoisomerases. J Mol Biol 300(2):353–362. https://doi.org/10.1006/jmbi.2000.3844
Remus D, Beuron F, Tolun G, Griffith JD, Morris EP, Diffley JF (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell 139(4):719–730. https://doi:S0092-8674(09)01303-8 [pii] https://doi.org/10.1016/j.cell.2009.https://doi.org/10.015
Salinas F, Benkovic SJ (2000) Characterization of bacteriophage T4-coordinated leading- and lagging-strand synthesis on a minicircle substrate. Proc Natl Acad Sci USA 97(13):7196–7201. https://doi.org/10.1073/pnas.97.13.7196
Schauer G, Finkelstein J, O'Donnell M (2017) In vitro assays for eukaryotic leading/lagging strand DNA replication. Bio Protoc 7(18). https://doi.org/10.21769/BioProtoc.2548
Simon AC, Zhou JC, Perera RL, van Deursen F, Evrin C, Ivanova ME, Kilkenny ML, Renault L, Kjaer S, Matak-Vinkovic D, Labib K, Costa A, Pellegrini L (2014) A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. Nature 510(7504):293–297. https://doi:nature13234 [pii] https://doi.org/10.1038/nature13234
Smith DJ, Whitehouse I (2012) Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature 483(7390):434–438. https://doi.org/10.1038/nature10895
Steitz TA (1999) DNA polymerases: structural diversity and common mechanisms. J Biol Chem 274(25):17395–17398. https://doi.org/10.1074/jbc.274.25.17395

Stukenberg PT, Turner J, O'Donnell M (1994) An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. Cell 78(5):877–887. https://doi:http://dx.doi.org/S0092-8674(94)90662-9

Sun J, Shi Y, Georgescu RE, Yuan Z, Chait BT, Li H, O'Donnell ME (2015) The architecture of a eukaryotic replisome. Nat Struct Mol Biol 22(12):976–982. https://doi.org/10.1038/nsmb.3113

Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV (2009) Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. Biol Direct 4:11. https://doi:1745-6150-4-11 [pii] https://doi.org/10.1186/1745-6150-4-11

Tougu K, Marians KJ (1996) The interaction between helicase and primase sets the replication fork clock. J Biol Chem 271(35):21398–21405

van Deursen F, Sengupta S, De Piccoli G, Sanchez-Diaz A, Labib K (2012) Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. EMBO J 31(9):2195–2206. https://doi:emboj201269 [pii] https://doi.org/10.1038/emboj.2012.69

Wasserman MR, Schauer GD, O'Donnell ME, Liu S (2019) Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. Cell 178(3):600–611 e616. https://doi.org/10.1016/j.cell.2019.06.032

Watase G, Takisawa H, Kanemaki MT (2012) Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. Curr Biol 22(4):343–349. https://doi.org/10.1016/j.cub.2012.01.023

Yang W, Woodgate R (2007) What a difference a decade makes: insights into translesion DNA synthesis. Proc Natl Acad Sci USA 104(40):15591–15598. https://doi:0704219104 [pii] https://doi.org/10.1073/pnas.0704219104

Yang J, Zhuang Z, Roccasecca RM, Trakselis MA, Benkovic SJ (2004) The dynamic processivity of the T4 DNA polymerase during replication. Proc Natl Acad Sci USA 101(22):8289–8294. https://doi.org/10.1073/pnas.0402625101

Yao NY, O'Donnell M (2012) The RFC clamp loader: structure and function. Subcell Biochem 62:259–279. https://doi.org/10.1007/978-94-007-4572-8_14

Yao NY, O'Donnell ME (2016) Evolution of replication machines. Crit Rev Biochem Mol Biol 51(3):135–149. https://doi.org/10.3109/10409238.2015.1125845

Yao NY, O'Donnell ME (2019) Machinery of DNA replication. In: Roberts G, Watts A (eds) Encyclopedia of biophysics. Springer, Berlin, Heidelberg. doi:https://doi.org/10.1007/978-3-642-35943-9

Yao NY, Georgescu RE, Finkelstein J, O'Donnell ME (2009) Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. Proc Natl Acad Sci USA 106(32):13236–13241. https://doi:0906157106 [pii] https://doi.org/10.1073/pnas.0906157106

Yu C, Gan H, Serra-Cardona A, Zhang L, Gan S, Sharma S, Johansson E, Chabes A, Xu RM, Zhang Z (2018) A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. Science 361(6409):1386–1389. https://doi.org/10.1126/science.aat8849

Yuan Z, Bai L, Sun J, Georgescu R, Liu J, O'Donnell ME, Li H (2016) Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. Nat Struct Mol Biol 23(3):217–224. https://doi.org/10.1038/nsmb.3170

Yuan Z, Georgescu R, Santos RLA, Zhang D, Bai L, Yao NY, Zhao G, O'Donnell ME, Li H (2019) Ctf4 organizes sister replisomes and Pol alpha into a replication factory. Elife 8. https://doi.org/10.7554/eLife.47405

Chapter 6
Recent Progress in Structural Studies on the GT-C Superfamily of Protein Glycosyltransferases

Heather Bohl, Lin Bai, and Huilin Li

Abstract Protein glycosylation is an essential covalent modification involved in protein secretion, stability, binding, folding, and activity. One or more sugars may be O-, N-, S-, or C-linked to specific amino acids by glycosyltransferases, which catalyze the transfer of these sugars from a phosphate-containing carrier molecule. Most glycosyltransferases are members of the GT-A, GT-B, or GT-C structural superfamilies. GT-C enzymes are integral membrane proteins that utilize a phospho-isoprenoid carrier for sugar transfer. To-date, two families of GT-Cs involved in protein glycosylation have been structurally characterized: the family represented by PglB, AglB, and Stt3, which catalyzes oligosaccharide transfer to Asn, and the family represented by Pmt1 and Pmt2, which catalyzes mannose transfer to Thr or Ser. This chapter reviews progress made over recent years on the structure and function of these two GT-C families.

Keywords Protein N-glycosylation · Protein O-mannosylation · Glycosyltransferases · Glycosyltransferase C fold (GT-C fold) · Cryo-EM · Structural biology

Introduction

Protein glycosylation is a protein modification in which one or more sugar residues are attached to amino acids, typically Asn, Thr, Ser, or Trp (Spiro 2002). Such modifications are present in all domains of life. Protein glycosylation is important

H. Bohl · H. Li (✉)
Structural Biology Program, Van Andel Institute, Grand Rapids, Michigan 49503, USA
e-mail: Huilin.Li@vai.org

H. Bohl
e-mail: bohlx031@umn.edu

L. Bai
Department of Biochemistry and Biophysics, School of Basic Medical Sciences, Peking University, 100083 Beijing, P. R. China
e-mail: lbai@bjmu.edu.cn

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_6

for protein folding, stability, activity, binding, and secretion, and the mutation of the enzymes involved in the process may result in severely deleterious phenotypes or diseases in humans (Helenius and Aebi 2004; Haeuptle and Hennet 2009).

Protein glycosylation is catalyzed by glycosyltransferases (GTs), which transfer the sugar from a donor to an acceptor. In protein glycosylation, the acceptor is either the protein or a sugar already attached to the protein. The sugar donors are usually activated nucleotide sugars or lipid phosphate-linked sugars (Lairson et al. 2008). In forming the glycosidic bond between sugar and acceptor, there are two types of glycosylation mechanisms, either inverting or retaining, which refer to the inversion or retention of the stereochemistry of the anomeric carbon of the transferred sugar (Lairson et al. 2008). Inverting glycosyltransferases perform a simple S_N2 type reaction: the enzyme catalyzes the nucleophilic attack of oxygen, nitrogen, carbon, or sulfur on the anomeric carbon with simultaneous cleavage of the bond to the phosphate-containing leaving group (Lairson et al. 2008; Chang et al. 2011). Retaining glycosyltransferases use an S_Ni-like reaction having an oxocarbenium-phosphate ion pair intermediate and an interaction between the leaving group and the nucleophile on the same face of the sugar (Hurtado-Guerrero and Davies 2012; Breton et al. 2012; Lairson et al. 2008; Yu et al. 2015; Lee et al. 2011). This mechanism is predicted for all retaining GTs (Hurtado-Guerrero and Davies 2012; Breton et al. 2012; Lairson et al. 2008).

Based on their structural fold, glycosyltransferases fall into one of three superfamilies: GT-A, GT-B, or GT-C (Lairson et al. 2008). GT-A enzymes form a domain composed of a seven-stranded β-sheet surrounded by α-helices and a small, two-stranded β-sheet that is sometimes described as two "closely abutted" Rossmann-like domains (Lairson et al. 2008; Chang et al. 2011). GT-B enzymes form two Rossmann-like domains, each with a 6- or 7-stranded β-sheet surrounded by α-helices; these domains are separated by a flexible linker with the active site at the cleft between the domains (Chang et al. 2011; Lairson et al. 2008). GT-As are typically metal-dependent, having a conserved DxD motif involved in divalent cation binding, while GT-Bs are typically metal-independent (Lairson et al. 2008; Chang et al. 2011). The metal or a positively charged residue stabilizes the negatively charged leaving group, which universally contains a phosphate (Lairson et al. 2008). The GT-C enzymes are all membrane proteins predicted to containing from 8 to 13 transmembrane helices (Lairson et al. 2008). The known GT-C enzymes all use lipid phosphate-linked sugar donors, which make sense considering their hydrophobic properties. All known GT-Cs use an inverting mechanism, while GT-As and GT-Bs can be either inverting or retaining (Lairson et al. 2008; Chang et al. 2011). Because the structural study of membrane proteins is difficult, GT-Cs have been less well characterized until recently.

Relative to the other two families, GT-C has far fewer members. For example, there are over 200 glycosyltransferases in humans, but GT-C accounts for only about 10%. GT-C enzymes are mainly involved in N-glycosylation, glypiation (addition of glycosylphosphatidylinositol; GPI anchor), O-mannosylation, and C-mannosylation of proteins. Some GT-C enzymes have been found to be involved in lipid modifications. GT-C members mainly belong to 14 families within the Carbohydrate

Active enZyme (CAZy) database. Until now, structural work has been reported for only three CAZy families: GT39, GT66, and GT83, which catalyze the protein O-mannosylation, protein N-glycosylaton, and lipid A glycosylation, respectively (Table 6.1). Among them, the enzymes that catalyze the transfer of an oligosaccharide to Asn in N-glycosylation are the most well studied. The structures of oligosaccharyltransferase (OST) from several species (bacterial PglB, archaeal AglB, the yeast OST complex, and human OST) have been determined (Lizak et al. 2011; Matsumoto et al. 2013; Wild et al. 2018; Bai et al. 2018; Ramirez et al. 2019). The structure of the yeast PMT complex that catalyzes the transfer of mannose to Ser/Thr (Bai et al. 2019) and the structure of the bacterial ArnT that catalyzes the attachment of the cationic sugar 4-amino-4-deoxy-L-arabinose to lipid A have also been published (Petrou et al. 2016). Most recently, the structure of a mycobacterial lipid glycosyltransferase, arabinofuranosyltransferase D (AftD), has been described (Tan et al. 2019). Interestingly, although these enzymes have totally different functions, their core structures share the same fold. This fact may indicate all GT-C enzymes share a common fold, as GT-A and GT-B members do, although this still needs to be confirmed.

This chapter focuses on advances in the structural understanding of the GT-Cs superfamily catalyzing protein glycosylation, over the past two years. We will introduce the structures of bacteria PglB, archaeal AglB, yeast and mammalian OSTs, and yeast PMT, and discuss their contribution to our understanding of the mechanisms and selectivity of protein glycosylation.

The Prokaryotic Protein *N*-glycosyltransferases: PglB and AglB

PglB and AglB are homologs of the catalytic subunit of the eukaryotic oligosaccharyltransferase Stt3. They are present in eubacteria and archaea, respectively, and they glycosylate Asn (Matsumoto et al. 2013; Lizak et al. 2011). While eukaryotic OST acts on the luminal face of the ER membrane, these prokaryotic homologs act on the periplasmic face of the plasma membrane (Helenius and Aebi 2004; Matsumoto et al. 2013; Lizak et al. 2011). Most structural studies have focused on PglB from *Campylobacter lari* (Lizak et al. 2011; Napiorkowska et al. 2018, 2017), so this protein will be the focus of this section, unless otherwise noted. Like other members of the GT-C superfamily, PglB is an integral membrane protein (Lizak et al. 2011; Liu and Mushegian 2003). Crystal structures of this enzyme revealed that it has thirteen transmembrane helices (TMHs) with elongated periplasmic loops between TMH1 and 2 (EL1) and TMH9 and 10 (EL5) as well as a globular C-terminal periplasmic domain (Fig. 6.1a) (Lizak et al. 2011). EL1 is primarily involved in binding the periplasmic domain, while EL5 is flexible and contributes to binding to the periplasmic domain and the donor and acceptor substrates. The flexibility of EL5 is likely important for substrate binding and product release. Correspondingly, the formation of a disulfide

Table 6.1 Major known GT-C fold glycosyltransferases

		Human	S. cerevisiae	Archaea	Bacteria	GT family	Structure
Protein glycosylation	O-mannosylation	POMT1, POMT2	PMT1-7		Pmt, Rv1002c	GT39	Y
		TMTC1, TMTC2, TMTC3, TMTC4				GT105	
	N-glycosylation	STT3A, STT3B	STT3	AglB	PglB	GT66	Y
	Glypiation	PIGB, PIGZ	GPI10, SMP3			GT22	
		PIGM	GPI14			GT50	
		PIGV	GPI18			GT76	
	C-mannosylation	DPY19L1, DPY19L2, DPY19L3, DPY19L4				GT98	
Lipid glycosylation	N-glycosylation donor synthesis	ALG6, ALG8	ALG6, ALG8			GT57	
		ALG3	ALG3			GT58	
		ALG10	ALG10			GT59	
		ALG9, ALG12	ALG9, ALG12			GT22	
	Others		FKS1, GSC2			GT48	
					EmbA/B/C	GT53	
					ArnT	GT83	Y
					Rv3792	GT85	
					AftD	GT53	Y

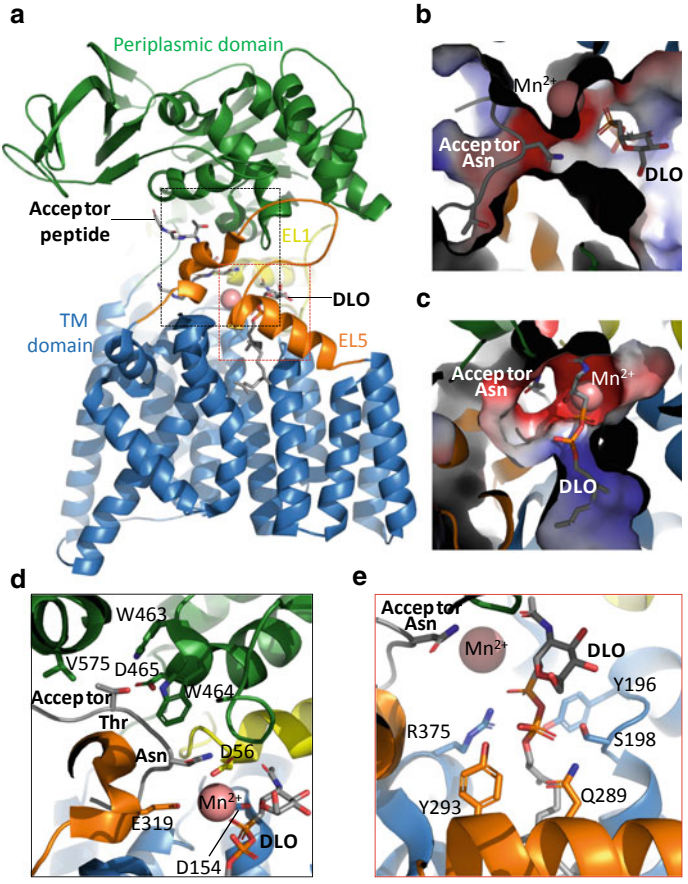


Fig. 6.1 Structure of the prokaryotic protein *N*-glycosyltransferase PglB. **a** PglB bound to an acceptor peptide and an inhibitory LLO (PDB ID 5OGL), colored by region: blue for the TM region, yellow for EL1, orange for EL5, and green for the periplasmic domain. Substrate carbons are gray and the Mn atom is pink. The areas in the dashed black and red rectangles are enlarged in panel D and E, respectively. **b, c** Electrostatic surface of the substrate binding site of PglB in top **b** and side **c** views. Binding of the acceptor Asn and the LLO on opposite sides of the, enzyme with red representing negative charge, blue representing positive charge, and white representing neutral. **d** Residues involved in recognition of the acceptor peptide NxT consensus site (PDB ID 5OGL). **e** Residues involved in binding the LLO (PDB ID 5OGL)

bond between EL5 and the transmembrane domain decreases activity for the natural lipid-linked oligosaccharide (LLO) substrate but not for a smaller synthetic LLO having only a single sugar.

PglB and other OSTs transfer an oligosaccharide from an isoprenoid lipid carrier to the nitrogen of an Asn side chain (Helenius and Aebi 2004). The acceptor peptide and donor LLO substrates bind between the periplasmic and transmembrane domains, and their binding pockets are connected by a small channel that the acceptor Asn

side chain pokes through (Fig. 6.1b, c) (Lizak et al. 2011; Napiorkowska et al. 2018, 2017). The catalytic residues reside on the LLO side of the enzyme while the residues that recognize the S/T of the NxS/T glycosylation consensus site (and, in bacteria, the D/E of the longer D/ExNxS/T consensus site) are located on the opposite side of the channel.

Crystal structures of PglB have been determined in complex with peptide only, peptide and inhibitory LLO, and LLO and inhibitory peptide substrates (Lizak et al. 2011; Napiorkowska et al. 2018, 2017). These structures revealed that the conserved WWD motif recognizes the hydroxyl of the acceptor S/T with three hydrogen bonds and that the Thr methyl has a further van der Waals interaction with Ile (Fig. 6.1d). In the active site, a divalent cation (here Mn^{2+}) is coordinated by Asp and Glu residues contributed by EL1, the short periplasmic loop EL2, and EL5 (Fig. 6.1a, d). At least one of these residues also interacts with the acceptor Asn (Fig. 6.1d). The LLO pyrophosphate is bound by the transmembrane domain with two contributions from EL5, and the *N*-acetylglucosamine (GlcNAc) at the base of the transferred oligosaccharide is bound by a conserved residue in the periplasmic domain and the metal-coordinating Asp residue from EL1 (Fig. 6.1e). In the catalytically competent binding conformation, the pyrophosphate may coordinate the divalent cation.

These structures provide clues about the catalytic mechanism. PglB and other OST catalytic subunits transfer the oligosaccharide with an inversion of the stereochemistry of the anomeric carbon bonded to the lipid carrier. Therefore, the Asn side chain nitrogen likely attacks the anomeric carbon by an S_N2-type mechanism (Lizak et al. 2011). However, the amide nitrogen is typically a poor nucleophile, because the nitrogen's lone pair is conjugated with the carbonyl and thus has pi bond character. One suggested mechanism involves twisting the amide to break this double-bond character and free the lone pair (Lizak et al. 2011). Another proposed mechanism focuses on the transformation of the anomeric carbon into a reactive electrophile via the electron-withdrawing action of the divalent cation, which is near (4.1 Å) the glycosidic oxygen in the structure of PglB bound to reactive LLO and inhibitory peptide (Napiorkowska et al. 2018). So far, no experimental evidence strongly favors either of these mechanisms.

Eukaryotic Protein *N*-glycosyltransferase: The OST Complexes

Unlike the prokaryotic OST that has a single catalytic subunit, the yeast OST is composed of eight integral membrane proteins (Wild et al. 2018; Bai et al. 2018). The yeast OST has two isoforms, Ost3, which associates with the Sec61 translocon, and Ost6, which associates with the Ssh1 translocon (Yan and Lennarz 2005). The structures of the *Saccharomyces cerevisiae* OST complex containing Ost3 has been determined by cryo-EM both solubilized in detergent (Bai et al. 2018) and reconstituted in nanodiscs (Wild et al. 2018) (Fig. 6.2a, b). The structures are quite similar,

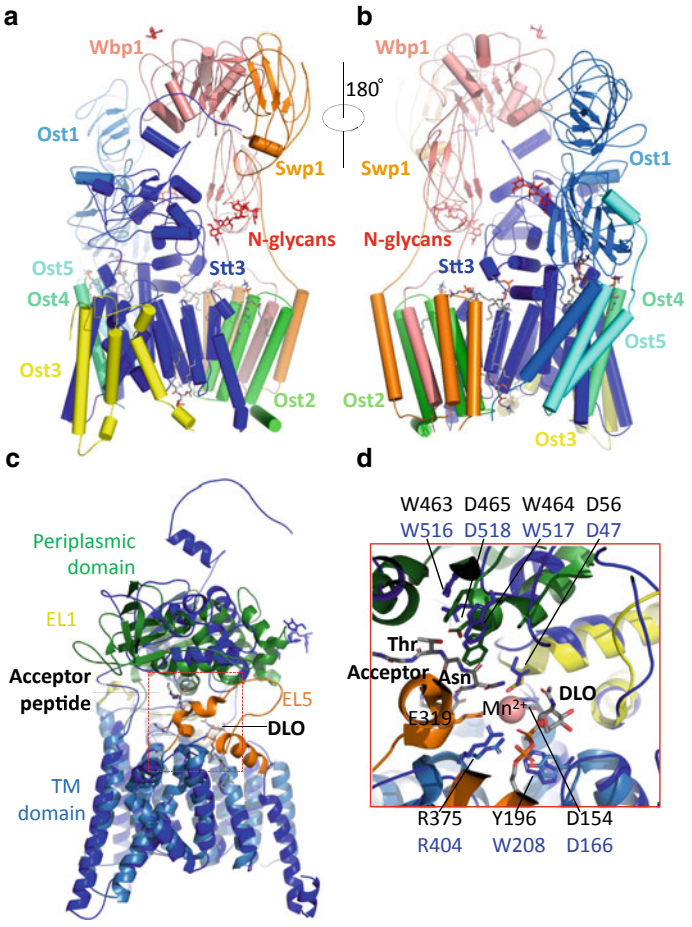


Fig. 6.2 Cryo-EM structure of the *S. cerevisiae* OST complex. The eight-protein complex is colored by subunit (PDB ID 6C26). Lipid and glycan carbons are gray and red, respectively. **a, b** Front and back views, with the luminal domains at the top. **c** Overlay of Stt3 (blue) and PglB (colored as in Fig. 6.1). **d** Residues involved in recognition of the acceptor peptide NxT and the donor DLO (PDB ID 5OGL and 6C26). Residues in blue are from Stt3, and those in black from PglB

with a root mean squared deviation (RMSD) of only 1.8 Å (Bai and Li 2019). In agreement with previous biochemical data, the structures showed that OST is divided into three subcomplexes that pack loosely in their transmembrane regions, and several ordered lipids were observed that appear to hold the OST complex together (Fig. 6.2a, b) (Bai et al. 2018; Wild et al. 2018; Mueller et al. 2015; Karaoglu et al. 1997). The three subcomplexes comprise Ost1 and Ost5; the catalytic subunit Stt3, Ost3, and Ost4; and Ost2, Wbp1, and Swp1 (Karaoglu et al. 1997; Mueller et al. 2015; Bai et al. 2018; Wild et al. 2018).

Although we know the atomic structures of the OST complex, the functions of many of the proteins are not well understood. The only protein with a well-established function is Stt3, the catalytic subunit. A comparison between PglB and Stt3 shows that the overall fold and catalytic residues are well conserved (Fig. 6.2c, d) (Wild et al. 2018; Bai et al. 2018), so their catalytic mechanisms are most likely the same. The luminal domains of Stt3, PglB, and AglB are more divergent than the transmembrane domains, with different insertions into a conserved core fold (Matsumoto et al. 2013).

We have clues about the roles of the other subunits, however, based on the structures of the yeast OST and of the mammalian OST in complex with the translocon and ribosome (Pfeffer et al. 2014; Braunger et al. 2018), as well as data from functional studies. Docking of the yeast OST structures into the cryo-ET map of the mammalian ribosome–translocon–OST supercomplex has indicated that Ost3 (or Ost6 in the other yeast isoform) directly interacts with the translocon, which agrees with previous data showing that Ost3 and Ost6 bind the Sec61 and Ssh1 translocons, respectively (Wild et al. 2018; Bai et al. 2018; Yan and Lennarz 2005). In addition, a phospholipid bound in a putative LLO-binding hydrophobic groove between Stt3 and Ost3 may indicate a role for Ost3 in LLO binding (Bai et al. 2018). The flexibility of Stt3 TMH9 and Ost3 TMH1, which were the only transmembrane helices not resolved in the complex, may also indicate a role in LLO binding (Bai et al. 2018; Wild et al. 2018). Finally, the luminal domain of Ost3, which was also not resolved, is known to be an oxidoreductase homologous to thioredoxin, and it is likely well positioned to help feed the nascent peptide from the Sec61 channel to the Stt3 active site (Bai et al. 2018; Schulz et al. 2009).

Other proteins in the complex appear to have primarily structural or scaffolding roles. Ost2 attaches the rest of its subcomplex to the transmembrane domain of Stt3 (Bai et al. 2018; Wild et al. 2018). Ost4 binds and likely stabilizes the transmembrane domain of Stt3, while Ost5 binds and likely stabilizes Ost1 (Wild et al. 2018; Bai et al. 2018). Other important elements for maintaining the integrity of the complex include the elongated C-terminal tail of Stt3 and the ordered glycan bonded to N539 of Stt3, both of which contact Wbp1 and Swp1 (Bai et al. 2018; Wild et al. 2018).

Finally, comparison to structures of known function allows speculation about the functions of the luminal domains of Ost1, Wbp1, and Swp1. On the basis of structural similarity to a noncatalytic domain of an aminopeptidase, the Ost1 luminal domains were proposed to capture the glycosylated peptide product to prevent reentry into the Stt3 active site (Bai et al. 2018). In addition, structural similarity suggested that the luminal domains of Wbp1 and Swp1 have a role in binding the LLO glycan (Bai et al. 2018). Further biochemical and structural studies are needed to determine the functions of these subunits and to elucidate the mechanisms of glycan binding and specificity.

Unlike yeast, mammals have two forms of the catalytic Stt3 subunit (STT3A and STT3B), leading to two OST complexes, OST-A and OST-B (Ruiz-Canada et al. 2009). OST-A binds the translocon and glycosylates peptides as they pass into the ER; OST-B acts as a proof-reader for NxS/T sites missed by OST-A (Ruiz-Canada et al. 2009; Shrimal et al. 2017; Cherepanova et al. 2014). A near-atomic-resolution structure of the mammalian OST-A bound to the Sec61 translocon and the ribosome

showed that OST docks to the translocon 6.5 nm from the Sec61 lateral gate that releases transmembrane helices into the ER membrane (Braunger et al. 2018). This explained why NxS/T sites are often missed when they are near transmembrane helices (Braunger et al. 2018; Nilsson and von Heijne 1993). In mammalian OST-A and OST-B, Ost3/6 are replaced by DC2 and KCP2 or by TUSC3 and MAGT1, respectively (Shrimal et al. 2017; Cherepanova et al. 2014). DC2 was found to mediate the binding of STT3A to the translocon, and the helices of STT3A that interfaced with DC2 are different in STT3B (Braunger et al. 2018). Moreover, ribophorin-I (homologous to yeast Ost1) binds the ribosome via a C-terminal cytosolic domain. The most recent cryo-EM structures of human OST-A and OST-B reveal that they are highly similar (RMSD of less than 1 Å). The ribosome-interacting C-terminal domain of ribophorin-I is an ordered four-helix bundle in isoform A, but the corresponding domain in isoform B is disordered, explaining why only OST-A binds to the ribosome for co-translational protein N-glycosylation (Ramirez et al. 2019).

Eukaryotic Protein *O*-mannosyltransferases: Pmt1 and Pmt2

Pmt proteins are GT-C superfamily glycosyltransferases that attach mannose to Ser or Thr (Loibl and Strahl 2013). Like OST, Pmt proteins act at the luminal face of the ER membrane and transfer mannose from a lipid carrier (Loibl and Strahl 2013; Helenius and Aebi 2004). Pmt proteins can be separated into three classes: PMT1, PMT2, and PMT4 (Loibl and Strahl 2013; Neubert and Strahl 2016). Animals encode only a single member each from the PMT2 and PMT4 classes and no PMT1 (Loibl and Strahl 2013). In yeast, PMT1 enzymes form heterodimers with PMT2 enzymes, and PMT4 enzymes homodimerize, but in animals PMT2 (POMT2) dimerizes with PMT4 (POMT1) (Loibl and Strahl 2013; Neubert and Strahl 2016). Recently, the atomic resolution structure of the heterodimer of Pmt1 and Pmt2 from *Saccharomyces* was solved by cryo-EM (Fig. 6.3a) (Bai et al. 2019).

The structures of Pmt1 and Pmt2 are highly similar, with their transmembrane domains overlaying with a RMSD of 1.8 Å and their lumenal MIR domains (domain present in mannosyltransferase, inositol triphosphate and ryanodine receptor) overlaying with a RMSD of 1.6 Å. The transmembrane domains each comprise 11 transmembrane helices with elongated luminal loops between TMH1 and 2 (LL1) and between TMH7 and 8 (LL4) (Fig. 6.3a). LL4 contains the MIR domain in each subunit, and these domains form a β-trefoil. The MIR domain of Pmt1 interacts with LL4 of Pmt2, but the MIR domain of Pmt2 was detached and was disordered in the cryo-EM map. The structure of this domain was solved by X-ray crystallography and docked into the cryo-EM map (Bai et al. 2019). Interactions between the subunits occur primarily in the cytosolic and luminal regions, leaving a large gap between Pmt1 and Pmt2 in the transmembrane region, which may allow facile diffusion of the lipid-linked mannose into the active sites. The attachment of the Pmt1 MIR domain to Pmt2 may block entry on this side. The functional importance of the MIR domains remains to be established, but they may bind to chaperones (Loibl and Strahl 2013).

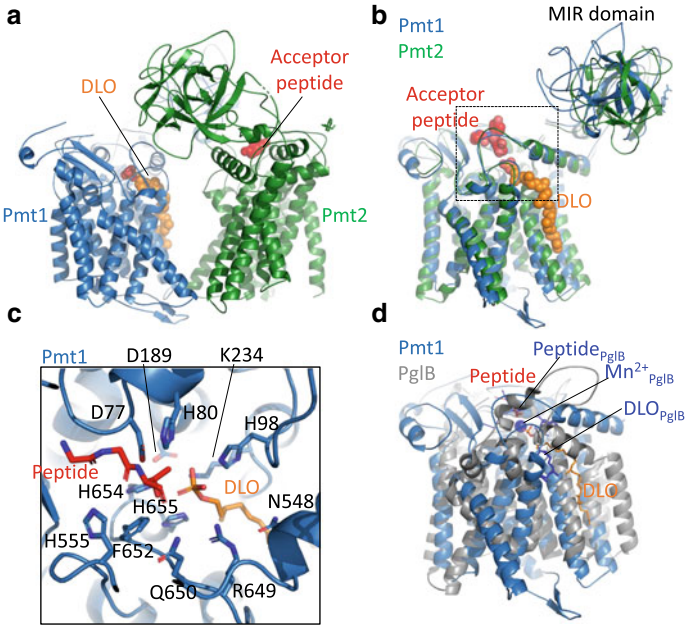


Fig. 6.3 Cryo-EM structure and the catalytic site of the *S. cerevisiae* Pmt1–Pmt2 complex. Subunits are colored by region. Pmt1 and Pmt2 are sky blue and forest green, respectively. Carbons of donor DLO and acceptor peptide are orange and red, respectively. **a** Model of the Pmt1–Pmt2 heterodimer (PDB ID 6P25). **b** Overlay of Pmt1 and Pmt2. The area in the dashed black rectangle are enlarged in panel C. **c** Residues involved in binding acceptor peptide and dol-P product. **d** Overlay of the Pmt1 and PglB TM regions

The structure of the heterodimer was solved with and without a tetrapeptide acceptor substrate (PYTV, where T is the glycosylated Thr) bound to Pmt2 (Bai et al. 2019). In both structures, an elongated electron density was associated with Pmt1 and modeled as dolichol phosphate (dol-P), the lipid carrier. Overlaying the structures of Pmt1 and Pmt2 allowed the ternary complex of either to be modeled and revealed the residues involved in binding the dol-P phosphate as well as the acceptor Thr (Fig. 6.3b, c). Importantly, an invariant DE motif was observed to be involved in acceptor binding, with the Asp positioned to activate the Thr nucleophile and Glu stabilizing the position of the motif by formation of a salt bridge with a conserved Arg.

The Pmt1 and Pmt2 structures also overlaid reasonably well with PglB and Stt3 (Fig. 6.3d), particularly in the transmembrane domains (RMSD of 3.2 Å between Pmt1 and PglB), even though PglB and Stt3 have two more transmembrane helices (Bai et al. 2019). In addition, the membrane proximal helices in EL1 of PglB are similar to those in LL1 of Pmt1 and Pmt2, and the membrane proximal helix in LL4 of Pmt1 and Pmt2 is similar to the helix in EL5 of PglB when LLO was bound. The MIR domains of Pmt1 and Pmt2 and the C-terminal periplasmic domain of PglB are

not similar. Comparison to PglB showed that the metal-binding residues of PglB are conserved in Pmt1/2, though no metal was observed bound to either subunit of the heterodimer. Moreover, the substrates of Pmt1/2 are bound similarly to those of PglB. Because Pmt1, Pmt2, and PglB are all inverting glycosyltransferases in the GT-C superfamily, the structural and mechanistic similarities are not unexpected. However, the nature of the Asn nucleophile of PglB may require specialized activation, as mentioned above (Lizak et al. 2011).

Summary

While structural information about the GT-C superfamily has long been lacking in comparison to the GT-A and GT-B superfamilies of glycosyltransferases, improvements in the crystallography of membrane proteins and recent improvements in the resolution routinely accessible by cryo-EM have begun to close that gap. PglB/AglB/Stt3 and Pmt1/2 represent two GT-C families (https://www.cazy.org). Determining the structures of GT-Cs from other families will allow us to identify elements that are conserved across all GT-Cs and elements specific to certain families. Understanding these similarities and differences may aid in the development of drugs specific to particular GT-Cs present in human pathogens or implicated in human diseases. The structures discussed here have greatly improved our understanding of the mechanism of protein glycosylation, though questions remain. In particular, the functions of numerous noncatalytic subunits and domains present in the OST complex have yet to be firmly established.

References

Bai L, Kovach A, You Q, Kenny A, Li H (2019) Structure of the eukaryotic protein O-mannosyltransferase Pmt1-Pmt2 complex. Nat Struct Mol Biol 26(8):704–711. https://doi.org/10.1038/s41594-019-0262-6

Bai L, Li H (2019) Cryo-EM is uncovering the mechanism of eukaryotic protein N-glycosylation. FEBS J 286(9):1638–1644. https://doi.org/10.1111/febs.14705

Bai L, Wang T, Zhao G, Kovach A, Li H (2018) The atomic structure of a eukaryotic oligosaccharyltransferase complex. Nature 555(7696):328–333. https://doi.org/10.1038/nature25755

Braunger K, Pfeffer S, Shrimal S, Gilmore R, Berninghausen O, Mandon EC, Becker T, Forster F, Beckmann R (2018) Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. Science 360(6385):215–219. https://doi.org/10.1126/science.aar7899

Breton C, Fournel-Gigleux S, Palcic MM (2012) Recent structures, evolution and mechanisms of glycosyltransferases. Curr Opin Struct Biol 22(5):540–549. https://doi.org/10.1016/j.sbi.2012.06.007

Chang A, Singh S, Phillips GN Jr, Thorson JS (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Curr Opin Biotechnol 22(6):800–808. https://doi.org/10.1016/j.copbio.2011.04.013

Cherepanova NA, Shrimal S, Gilmore R (2014) Oxidoreductase activity is necessary for N-glycosylation of cysteine-proximal acceptor sites in glycoproteins. J Cell Biol 206(4):525–539. https://doi.org/10.1083/jcb.201404083

Haeuptle MA, Hennet T (2009) Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. Hum Mutat 30(12):1628–1641. https://doi.org/10.1002/humu.21126

Helenius A, Aebi M (2004) Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem 73:1019–1049. https://doi.org/10.1146/annurev.biochem.73.011303.073752

Hurtado-Guerrero R, Davies GJ (2012) Recent structural and mechanistic insights into post-translational enzymatic glycosylation. Curr Opin Chem Biol 16(5–6):479–487. https://doi.org/10.1016/j.cbpa.2012.10.013

Karaoglu D, Kelleher DJ, Gilmore R (1997) The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. J Biol Chem 272(51):32513–32520. https://doi.org/10.1074/jbc.272.51.32513

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77:521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322

Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG (2011) Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. Nat Chem Biol 7(9):631–638. https://doi.org/10.1038/nchembio.628

Liu J, Mushegian A (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci 12(7):1418–1431. https://doi.org/10.1110/ps.0302103

Lizak C, Gerber S, Numao S, Aebi M, Locher KP (2011) X-ray structure of a bacterial oligosaccharyltransferase. Nature 474(7351):350–355. https://doi.org/10.1038/nature10151

Loibl M, Strahl S (2013) Protein O-mannosylation: what we have learned from baker's yeast. Biochim Biophys Acta 1833 (11):2438–2446. https://doi.org/10.1016/j.bbamcr.2013.02.008

Matsumoto S, Shimada A, Nyirenda J, Igura M, Kawano Y, Kohda D (2013) Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. Proc Natl Acad Sci U S A 110(44):17868–17873. https://doi.org/10.1073/pnas.1309777110

Mueller S, Wahlander A, Selevsek N, Otto C, Ngwa EM, Poljak K, Frey AD, Aebi M, Gauss R (2015) Protein degradation corrects for imbalanced subunit stoichiometry in OST complex assembly. Mol Biol Cell 26(14):2596–2608. https://doi.org/10.1091/mbc.E15-03-0168

Napiorkowska M, Boilevin J, Darbre T, Reymond JL, Locher KP (2018) Structure of bacterial oligosaccharyltransferase PglB bound to a reactive LLO and an inhibitory peptide. Sci Rep 8(1):16297. https://doi.org/10.1038/s41598-018-34534-0

Napiorkowska M, Boilevin J, Sovdat T, Darbre T, Reymond JL, Aebi M, Locher KP (2017) Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. Nat Struct Mol Biol 24(12):1100–1106. https://doi.org/10.1038/nsmb.3491

Neubert P, Strahl S (2016) Protein O-mannosylation in the early secretory pathway. Curr Opin Cell Biol 41:100–108. https://doi.org/10.1016/j.ceb.2016.04.010

Nilsson IM, von Heijne G (1993) Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J Biol Chem 268(8):5798–5801

Petrou VI, Herrera CM, Schultz KM, Clarke OB, Vendome J, Tomasek D, Banerjee S, Rajashankar KR, Belcher Dufrisne M, Kloss B, Kloppmann E, Rost B, Klug CS, Trent MS, Shapiro L, Mancia F (2016) Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. Science 351(6273):608–612. https://doi.org/10.1126/science.aad1172

Pfeffer S, Dudek J, Gogala M, Schorr S, Linxweiler J, Lang S, Becker T, Beckmann R, Zimmermann R, Forster F (2014) Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. Nat Commun 5:3072. https://doi.org/10.1038/ncomms4072

Ramirez AS, Kowal J, Locher KP (2019) Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. Science 366(6471):1372–1375. https://doi.org/10.1126/science.aaz3505

Ruiz-Canada C, Kelleher DJ, Gilmore R (2009) Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. Cell 136(2):272–283. https://doi.org/10.1016/j.cell.2008.11.047

Schulz BL, Stirnimann CU, Grimshaw JP, Brozzo MS, Fritsch F, Mohorko E, Capitani G, Glockshuber R, Grutter MG, Aebi M (2009) Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. Proc Natl Acad Sci U S A 106(27):11061–11066. https://doi.org/10.1073/pnas.0812515106

Shrimal S, Cherepanova NA, Gilmore R (2017) DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon. J Cell Biol 216(11):3625–3638. https://doi.org/10.1083/jcb.201702159

Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12(4):43r–56r. https://doi.org/10.1093/glycob/12.4.43r

Tan YZ, Zhang L, Rodrigues J, Zheng RB, Giacometti SI, Rosário AL, Kloss B, Dandey VP, Wei H, Brunton R, Raczkowski AM, Athayde D, Catalão MJ, Pimentel M, Clarke OB, Lowary TL, Archer M, Niederweis M, Potter CS, Carragher B, Mancia F (2019) Cryo-EM structures and regulation of arabinofuranosyltransferase AftD from Mycobacteria. bioRxiv

Wild R, Kowal J, Eyring J, Ngwa EM, Aebi M, Locher KP (2018) Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. Science 359(6375):545–550. https://doi.org/10.1126/science.aar5140

Yan A, Lennarz WJ (2005) Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. Glycobiology 15(12):1407–1415. https://doi.org/10.1093/glycob/cwj026

Yu H, Takeuchi M, LeBarron J, Kantharia J, London E, Bakker H, Haltiwanger RS, Li H, Takeuchi H (2015) Notch-modifying xylosyltransferase structures support an SNi-like retaining mechanism. Nat Chem Biol 11(11):847–854. https://doi.org/10.1038/nchembio.1927

Chapter 7
How Structures of Complement Complexes Guide Therapeutic Design

Jasmine K. Bickel, Tomas B. Voisin, Edward W. Tate, and Doryen Bubeck

Abstract The complement system is essential for immune defence against infection and modulation of proinflammatory responses. Activation of the terminal pathway of complement triggers formation of the membrane attack complex (MAC), a multi-protein pore that punctures membranes. Recent advances in structural biology, specifically cryo-electron microscopy (cryoEM), have provided atomic resolution snapshots along the pore formation pathway. These structures have revealed dramatic conformational rearrangements that enable assembly and membrane rupture. Here we review the structural basis for MAC formation and show how soluble proteins transition into a giant β-barrel pore. We also discuss regulatory complexes of the terminal pathway and their impact on structure-guided drug discovery of complement therapeutics.

Keywords Membrane attack complex · Complement therapeutics · Structure-guided drug discovery · Cryo electron microscopy · MACPF · C5

J. K. Bickel · T. B. Voisin · D. Bubeck (✉)
Department of Life Sciences, Imperial College London, Sir Ernst Chain Building, South Kensington Campus, London SW7 2AZ, UK
e-mail: d.bubeck@imperial.ac.uk

J. K. Bickel
e-mail: j.bickel15@imperial.ac.uk

T. B. Voisin
e-mail: tomas.voisin15@imperial.ac.uk

J. K. Bickel · E. W. Tate
Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, White City Campus, London W12 0BZ, UK
e-mail: e.tate@imperial.ac.uk

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_7

Complement

Complement is a protein-based network that is a fundamental component of immune defence. This complex system involves over 30 plasma and membrane-associated serum proteins. Complement proteins are involved in all stages of immune clearance from initial pathogen detection to elimination of infection (Fig. 7.1) (Ricklin et al. 2010). A complement response can be initiated either by recognizing the surface-bound sugars of pathogens, through the lectin pathway, or the antigens presented on cell surfaces, triggering the classical pathway. By contrast, the alternative pathway of complement is initiated by spontaneous hydrolysis of the serum protein C3 to form C3b, a process referred to as tick-over (Pangburn et al. 1981; Bexborn et al. 2008). Broadly, complement activation triggers proteolytic cleavage events that irreversibly propagate a series of cascades. If left unchecked, these cascades cause proinflammatory responses, phagocytosis of microbes, and lysis of targeted cells (Ricklin et al. 2016). Complement also modulates trafficking of foreign antigens for clearance by T- and B-cell lymphocytes, providing a functional bridge between adaptive and innate immune responses (Carroll 2004; Dunkelberger and Song 2010).

The terminal pathway is the direct killing arm of complement and can be activated by any of the upstream cascades (Fig. 7.1). The lectin, classical, and alternative pathways generate a complex that cleaves complement protein C5, referred to as the C5

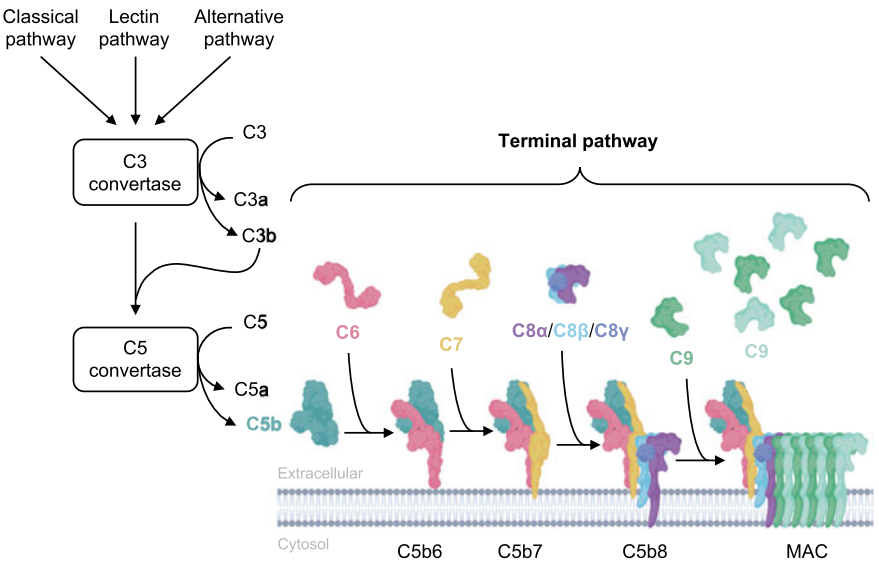


Fig. 7.1 Schematic outlining how the three upstream complement cascades (classical pathway, lectin pathway, and alternative pathway) activate the terminal pathway to generate the MAC pore. MAC is formed on target cell membranes by the sequential assembly of C5b, C6, C7, C8 and several copies of C9. Intermediate subcomplexes (C5b6, C5b7, C5b8) and a cross-section through the final pore are shown

convertase. The proteolytic activity of the C5 convertase is responsible for the irreversible assembly of the membrane attack complex (MAC), a multi-protein immune pore that ruptures membranes (Bubeck 2014; Bayly-Jones et al. 2017). MAC directly lyses and kills a multitude of pathogens, including enveloped viruses, parasites, and Gram-negative bacteria (Hoover et al. 1984; Tomlinson et al. 1989; Nakamura et al. 1996). Furthermore, the pore is essential for defence against *Neisseria meningitidis*, with genetic deficiencies in terminal complement proteins leading to recurrent meningococcal and gonococcal infections (Harriman et al. 1981; Schneider et al. 2007; Botto et al. 2009).

Beyond lysis, deposition of MAC pores on the surface of self-cells impacts diverse cellular processes. These pores, referred to as sub-lytic, cause an influx of calcium ions (Ca^{2+}) into the cell and activate downstream effector functions ranging from proliferation to apoptosis (Morgan 1989; Xie et al. 2020). Ion influx can also induce proinflammatory responses (Morgan 2016), resulting in secretion of cytokines such as interleukin (IL)-6, -8, and the vascular endothelial growth factor (VEGF) (Lueck et al. 2011). The effects of sub-lytic MAC include further downstream inflammatory effectors and initiate NLRP3 inflammasome assembly (Triantafilou et al. 2013). While many of these interaction partners have yet to be identified, a signalling complex within Rab5(+) endosomes has been shown to activate non-canonical NF-κB in endothelial cells (Jane-wit et al. 2015; Fang et al. 2019).

The balance between activation and regulation of complement is essential for tissue homeostasis during an immune response. Understanding the assembly and regulation of MAC at a molecular level will likely underpin the development of new therapeutics that tune complement activity. In this chapter we will review individual terminal pathway proteins and how they assemble into a lytic pore on target cells. We will also investigate the structural basis for how this process is controlled, starting from initiation of the terminal pathway by the C5 convertase to catching MAC pores on the membrane at the final step in the cascade. Finally, we will discuss how structural studies of complement complexes can support drug discovery.

Understanding MAC at the Molecular Level

The "Resolution Revolution" of cryo-electron microscopy (cryoEM) has swept across the complement field, providing scientists and clinicians an unprecedented view into how these complex machines function. The first electron microscopy images of the MAC were taken in the early 1980s and in agreement with electrophysiology experiments, these images revealed the formation of a large ring-shaped pore that punched holes in target membranes (Podack et al. 1980). However, crucial details about how individual blood-based proteins convert to form a transmembrane pore remained unanswered. In the last few years, cryoEM has transformed the field of structural biology, with the biggest step change coming from improvements in image detection that allow direct recording of electrons (Faruqi and Henderson 2007; Li et al. 2013; Wu et al. 2016). New computer algorithms have exploited the better

signal-to-noise ratio of the collected images, dramatically improving the accuracy of aligning individual single particles (Punjani et al. 2017; Zivanov et al. 2018). As a result, multiple conformations of complexes can be disentangled to provide not just one structure, but a portfolio of snapshots illustrating machines in action. This type of in silico purification is essential when trying to visualise complexes where further biochemical purification is simply not possible. Further adding to the complexity of preparing homogeneous samples for structural studies, large macromolecular complexes are often extremely flexible. The inherent mobility of these machines limits their ability to be captured within the restraints of a crystallographic lattice, an obligate requirement for X-ray diffraction experiments. With its ability to visualise large heterogeneous complexes at atomic resolution, cryoEM has revealed key insights into both the assembly and regulation of the complement MAC pore (Dudkina et al. 2016; Sharp et al. 2016; Serna et al. 2016; Menny et al. 2018; Spicer et al. 2018).

Molecular Assembly of MAC

Activation of the terminal pathway initiates an irreversible signalling cascade that ends in the formation of MAC (Fig. 7.1). The C5 convertase triggers MAC assembly on target cells by proteolytically cleaving C5 into two fragments: C5a and C5b (DiScipio et al. 1983). While C5a is a small anaphylatoxin that goes on to activate inflammatory pathways through GPCR signalling (Gerard and Gerard 1991; Guo and Ward 2005), C5b serves as a platform for assembly of complement proteins C6, C7, C8 and C9 (Podack 1984). The proteolytic cleavage of C5 causes a dramatic conformational re-arrangement of its domain architecture (Fig. 7.2a, b) (Hadders et al. 2012). The thioester domain (TED) of C5 is released like a coiled spring to extend half-way along a scaffold of macroglobulin (MG) domains. C6 captures the newly cleaved C5b to form the C5b6 complex (Fig. 7.2c) (Cooper and Müller-Eberhard 1970; Hadders et al. 2012). In the absence of C6, this intermediate C5b conformation is highly labile and will decay to a state unable to continue the cascade (Cooper and Müller-Eberhard 1970). In plasma, C6 is an elongated molecule whereby its core Membrane Attack Complex Perforin Fold (MACPF) domain is sandwiched by a series of smaller auxiliary domains tethered by flexible linkers (Aleshin et al. 2012). The crystal structure of C5b6 revealed that the C-terminal domains of C6, together with a flexible linker, trap the transient position of the C5 TED domain and are essential for initiating MAC formation (Hadders et al. 2012). C7 is the next protein to join the growing MAC precursor. Although there is currently no structural information for soluble C7, the protein is thought to undergo dramatic conformational rearrangements upon membrane insertion. C7 displaces the assembly precursor from the convertase and results in irreversible interactions with the membrane (Preissner et al. 1985; DiScipio et al. 1988). The hetero-trimeric protein C8, comprised of C8α C8β and C8γ, is subsequently recruited. Large-scale conformational changes in C8 induced by binding the MAC precursor enable C8β to associate with the newly

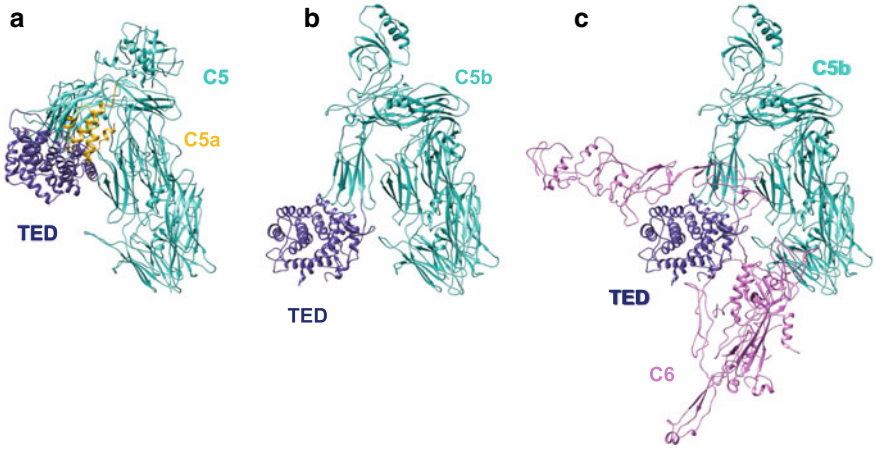


Fig. 7.2 Conformational changes of C5 initiate MAC assembly. **a** Ribbon diagram of the C5 crystal structure (PDB 3CU7). C5a is shown in gold, the TED domain in dark blue and the remainder of the structure in turquoise. **b** Ribbon diagram showing the conformation of C5b within the C5b6 complex (PDB 4A5W). Following cleavage of C5a, the C5b TED domain (dark blue) moves along the MG scaffold. **c** Structure of the C5b6 complex. Binding of C6 (magenta) holds the C5b TED domain (dark blue) in position. Remainder of C5b is in turquoise

exposed face of C7 (Stewart et al. 1987; Brannen and Sodetz 2007). In addition, C8γ rotates from its position in the soluble complex to reveal the nascent C9 binding site on C8α (Serna et al. 2016). C9 is the last component of the cascade and homo-oligomerizes to complete the final pore (Podack et al. 1982; Tschopp et al. 1985; Parsons et al. 2019).

Interactions with the Membrane

Complement proteins interact with the target membrane at each stage of MAC assembly. The C5b6 complex, though soluble, can bind reversibly to lipid bilayers (Silversmith and Nelsestuen 1986; Parsons et al. 2019). Biophysical experiments have shown that C5b6 preferentially binds negatively charged lipids (Yorulmaz et al. 2015; Parsons et al. 2019). This interaction reduces the energy required to bend the lipid bilayer and primes the membrane for rupture (Menny et al. 2018). Based on current structural data, it remains unclear which residues of C5b6 are initially responsible for membrane binding, however, an α-helix with amphipathic characteristics in the C6 thrombospondin domain 1 (TS1) is a potential candidate (Aleshin et al. 2012; Hadders et al. 2012). Binding of C7 to membrane-bound C5b6 triggers the irreversible association of the complex with lipids, while C8 is the first component to pierce the bilayer (Podack et al. 1982; Steckel et al. 1983). Incorporation of C8 renders the membrane permeable to Ca^{2+} ions (Campbell et al. 1981; Morgan 1989),

which in turn causes the membrane to stiffen (Menny et al. 2018). The addition of C9 into the assembling MAC results in further rigidification of the bilayer as the pore widens (Menny et al. 2018).

MAC is comprised of an asymmetric and irregular β-barrel pore (Sharp et al. 2016; Serna et al. 2016; Menny et al. 2018). Unlike related homo-oligomeric pores, MAC is a hetero-oligomeric complex whereby each complement protein contributes two β-hairpins (Menny et al. 2018). The length and sequence of these contributing β-hairpins differ across the complex. Hairpins of early assembly precursors skate across the outer leaflet of the lipid bilayer and are anchored by aromatic residues at the tips. A positively charged patch above the hairpin tips likely interacts with lipid headgroups and distorts the planarity of the bilayer. By contrast, contributing hairpins of C8 and C9 are longer and lack the nonplanar charged patch exhibited by earlier components of the assembly (Menny et al. 2018). Residues of C8 and C9 that face the lipidic edge form a hydrophobic band the width of a bilayer, consistent with a transmembrane pore.

The principal pore-forming component of MAC lies within its MACPF domain. With the exception of C5b and C8γ, all MAC proteins possess a MACPF domain (Hadders et al. 2007; Lovelace et al. 2011; Aleshin et al. 2012; Spicer et al. 2018). The fold is characterized by a central kinked β-sheet, sandwiched by two helical clusters (Fig. 7.3) (Hadders et al. 2007; Rosado et al. 2007). Early crystallographic studies of complement proteins and bacterial pore-forming toxins of the cholesterol-dependent cytolysin (CDC) superfamily revealed that despite no obvious sequence similarity, their structures were highly conserved (Fig. 7.3) (Rossjohn et al. 1997; Polekhina et al. 2005; Hadders et al. 2007; Rosado et al. 2007; Law et al. 2010; Lovelace et al. 2011; Aleshin et al. 2012). During pore formation, the MACPF domain of complement proteins undergoes dramatic changes in secondary structure. Residues comprising one of the helical clusters unfurl to form two transmembrane hairpins (TMH) (Shepard et al. 1998; Shatursky et al. 1999; Rosado et al. 2007). Oligomerization of complement proteins during MAC assembly results in the alignment of these hairpins to form a large β-barrel pore (Fig. 7.4) (Hadders et al. 2012; Serna et al. 2016; Menny et al. 2018).

Structure of MAC

MAC is a multiprotein hetero-oligomeric pore formed from the assembly of MACPF domains. Taking advantage of recent advances in cryoEM, the MAC structure revealed the composition of the oligomer for the first time (Serna et al. 2016). The structure showed that the assembly precursor is comprised of a single copy of C5b6, C7 and C8, forming an integral component of the pore. The pore is completed by the addition of 18 copies of C9, which homo-oligomerize in the membrane (Fig. 7.4). Unlike other previously reported homo-oligomeric pores (Law et al. 2010; Dudkina et al. 2016), the MAC β-barrel is not symmetric (Sharp et al. 2016; Serna et al. 2016).

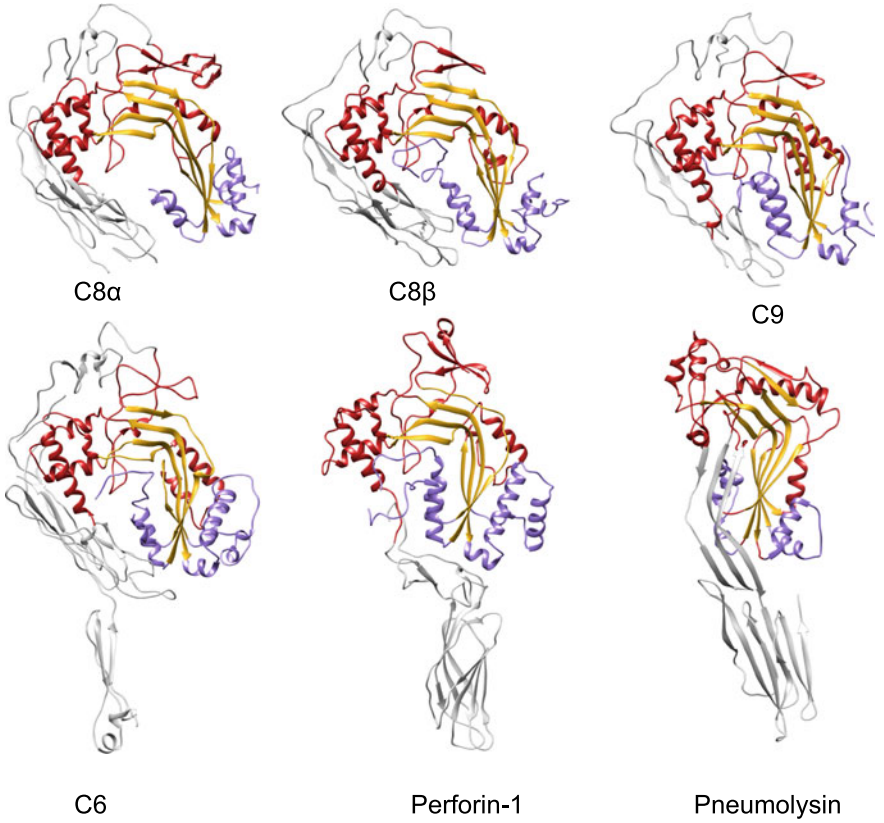


Fig. 7.3 Structural similarities between MACPF/CDC-containing proteins: C8α (PDB 3OJY), C8β (PDB 3OJY), C9 (PDB 6CXO), C6 (PDB 3T5O), Perforin-1 (PDB 3NSJ), and Pneumolysin (PDB 5AOD). MACPF/CDC domain (coloured) contains a bent β-sheet (gold) surrounded by helical clusters (red, purple). Residues that unfurl into transmembrane β-hairpins are purple. All other domains are in grey. Complement control protein (CCP)1, CCP2, TS2, TS3, Factor I membrane attack complex (FM)1 and FM2 domains of C6 are not shown for clarity

Instead, complement proteins are arranged in an asymmetric 'split-washer' configuration whereby the seam of the oligomer is not fully closed (Fig. 7.4b) (Serna et al. 2016). The unusual 'split-washer' architecture observed for detergent-solubilized MAC pores was confirmed by cryo-electron tomography studies of MAC embedded in a lipid model membrane (Sharp et al. 2016).

The interfaces between MACPF domains allow for a flexible assembly of proteins. This conformational variability of MAC accommodates variations in curvature and limited the resolution of these early structures. Using computational approaches that separated stoichiometrically equivalent states, the recent cryoEM MAC structure identified two distinct conformations, 'open' and 'closed' (Fig. 7.4) (Menny et al. 2018). Both states exhibit the characteristic 'split-washer' configuration of MAC,

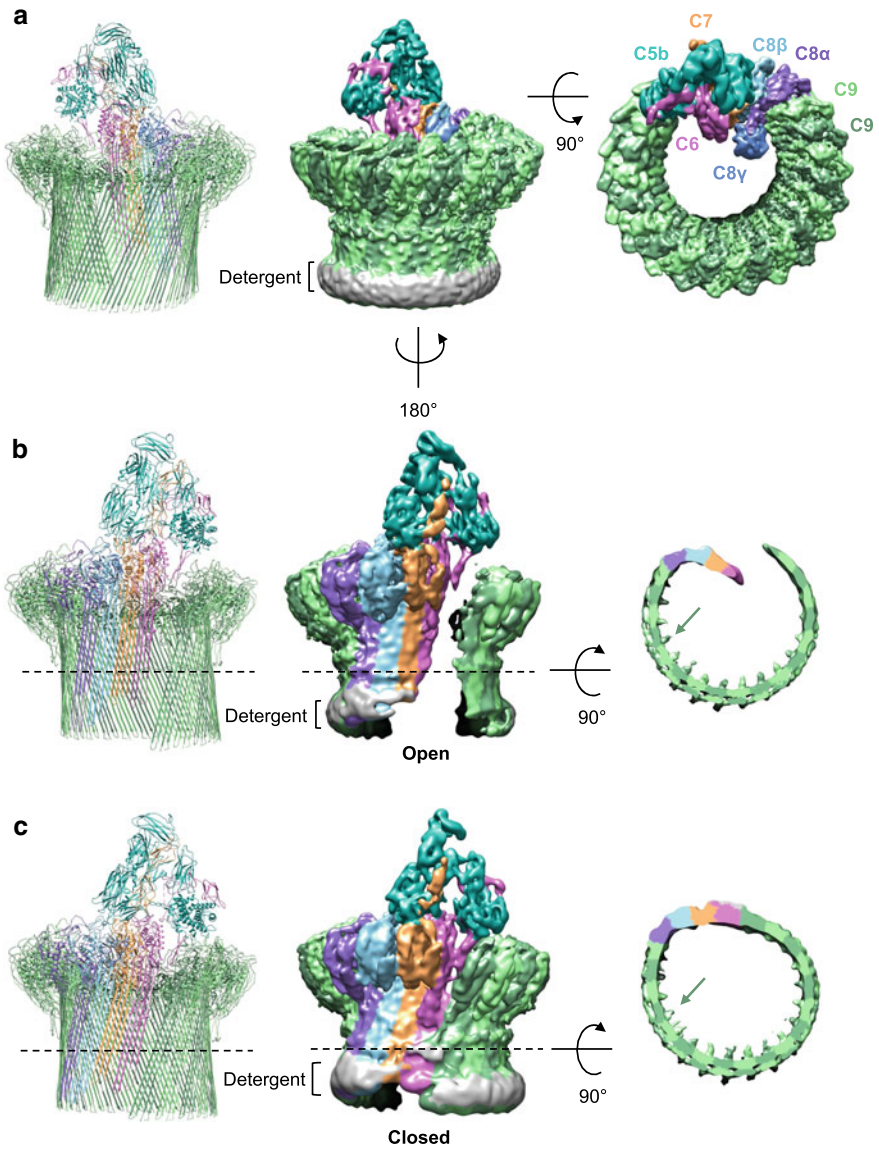


Fig. 7.4 CryoEM structure of MAC. **a**–**b** Ribbon diagram representation (PDB 6H03) and cryoEM reconstruction (EMD-0106) of the MAC 'open' conformation. Map and models in panels **a** and **b** are related by a 180° rotation. **c** Ribbon diagram (PDB 6H04) and cryoEM density map (EMD-0107) of the 'closed' conformation, orientation as in **b**. Dashed line in **b** and **c** indicate the position of the cross-section shown in the far-right panels. Green arrows indicate density for a C9 glycan. Maps and models are coloured according to protein composition, with the detergent belt indicated

however the central β-barrel is sealed in only the closed form (Fig. 7.4b, c). Further investigation is required to understand if these states are specific endpoints or exist in a dynamic equilibrium. The inherent flexibility of the MAC may be important for its ability to navigate complex target membrane environments. The surfaces of Gram-negative bacteria are coated with lipopolysaccharides and their outer membranes are densely packed with porins (Konovalova et al. 2017). An assembly that can adopt flexible curvatures could increase its efficiency for making large lesions; the MAC pore is over 100 Å in diameter (Serna et al. 2016; Menny et al. 2018). Furthermore, interconversion between the open and closed forms may add additional mechanical strain on the lipid bilayer, contributing to membrane destabilization.

Amidst this flexibility, MAC maintains its giant β-barrel through a number of core structural features. The central MACPF domains of complement proteins provide a major interaction interface for the complex and a sound foundation for the assembly (Hadders et al. 2012). However, the most extensive interfaces are located nearly 100 Å above the lipid bilayer. Along the barrel there are three key elements that prevent collapse of the β-barrel in the absence of large buried surface area (Menny et al. 2018). Firstly, The C-terminus of C9 wraps across the newly extended hairpins of an adjacent monomer. The collective C-termini then form a belt around the outer surface of the barrel. Juxtaposing the belt on the lumen of the barrel are concentric rings of glycans that provide shape and stability to an otherwise unsupported structure (Fig. 7.4c). Finally, disulfide bonds within β-hairpins further ensure that elongated strands maintain their secondary structure.

Converting Soluble Proteins into a Transmembrane Pore

MAC assembly requires large scale structural rearrangements of soluble proteins to form a transmembrane pore. In addition to the core pore-forming MACPF domain, complement proteins possess a number of auxiliary domains that ensure activation happens at the right place and time. Conserved across all MACPF-containing MAC proteins (C6, C7, C8α, C8β and C9), are core regulatory modules that include epidermal growth factor (EGF) and thrombospondin (TS) domains (Fig. 7.5) (Menny et al. 2018). Upon incorporation into MAC, EGF domains of complement proteins shift away from the helical bundles responsible for forming transmembrane hairpins (Menny et al. 2018). Conformational changes in both EGF and TS domains expose interfaces primed for binding the next component of the assembly. Likewise, a helix-turn-helix (HTH) in the MACPF which does not have a direct role in membrane insertion is displaced by adjacent monomers (Menny et al. 2018). C8γ has a similar role in the heterotrimeric C8 complex. Compared to its soluble form, C8γ rotates about an extended hairpin to expose the C9 binding site (Serna et al. 2016; Menny et al. 2018). While not essential for pore formation, C8γ enhances the efficiency of MAC assembly (Parker and Sodetz 2002), consistent with its role in regulating accessibility for C9. Conformational changes of auxiliary domains reveal

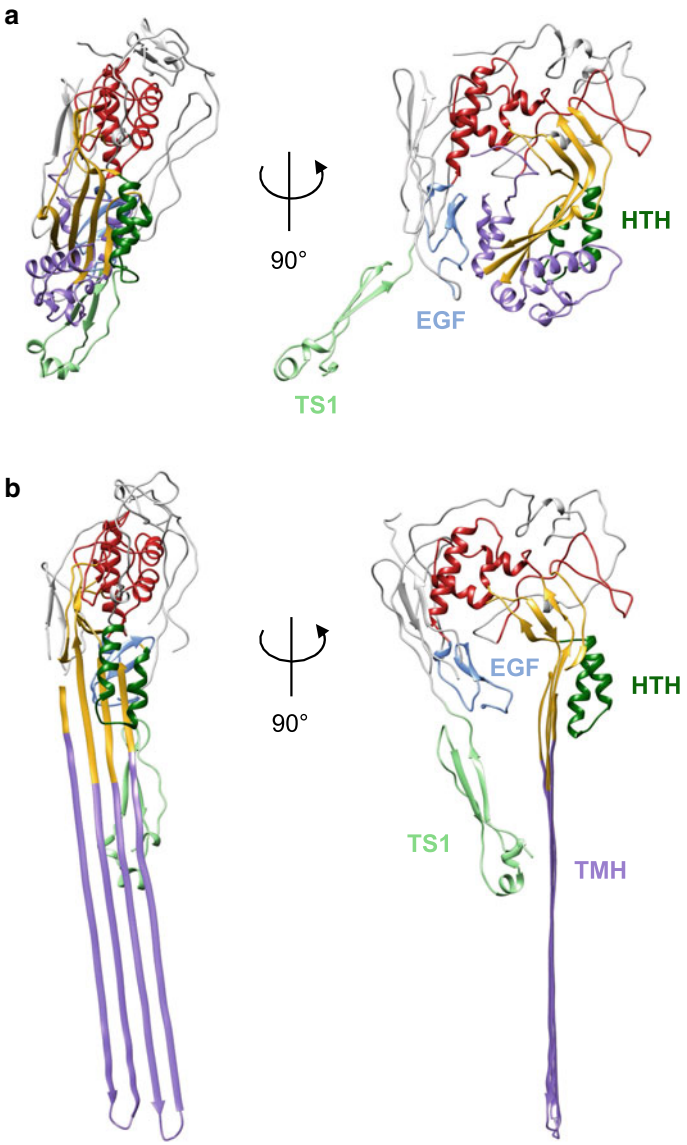


Fig. 7.5 Conformational changes in C6 upon MAC formation. **a** Ribbon diagram of soluble C6 (PDB 3T5O). **b** Ribbon diagram showing the conformation of C6 within the MAC 'open' conformation (PDB 6H03). Conformational rearrangements of the thrombospondin (TS) 1; green, epidermal growth factor (EFG; blue) and MACPF domains are highlighted. Within the MACPF: central β-sheet is gold, helical cluster is red, helix-turn-helix motif (HTH) is dark green, residues that form transmembrane hairpins (TMH) are purple. Low-density lipoprotein receptor class A, TS2 and TS3 are grey. All other domains have been removed for clarity

specific and highly charged surfaces on the leading MACPF edge. Oligomerization of MAC is driven largely by complementarity of electrostatic surface potentials between the newly exposed leading edge and the already accessible face of the next soluble complement protein (Menny et al. 2018). While there are currently no structures available of complexes stalled during assembly, C9 mutants that prevent the helix-to-hairpin transition of transmembrane residues also prevent further C9 oligomerization (Spicer et al. 2018).

Regulating MAC

During an immune response complement activation goes into overdrive. Unlike related bacterial homologues (CDCs), there is no lipid specificity for MAC pore formation. Once the terminal pathway is activated on a cell surface, any lipid bilayer is susceptible to rupture. As such, there are two main checkpoints for blocking MAC. The first prevents initiation of the cascade by inhibiting C5 cleavage. The second blocks membrane insertion at the final stages. Off-target assembly complexes are also scavenged and cleared by blood-based chaperones. Dysfunction in either of these control pathways can have devastating consequences for human health.

Overwhelming intrinsic control mechanisms for upstream complement pathways has knock-on consequences for MAC-mediated cell damage. Atypical haemolytic uremic syndrome (aHUS) is a disease characterized by acute renal failure and haemolytic anaemia (Noris and Remuzzi 2009). The pathogenesis of aHUS is heavily affected by uncontrolled amplification of the complement alternative pathway, resulting in overproduction of the C5 convertase on cell membranes (Noris and Remuzzi 2009; Roumenina et al. 2011). Consequently, MAC inserts into renal capillary endothelial cells causing damage, while C5a triggers proinflammatory signals in nearby immune cells. Similarly, age-related macular degeneration (AMD) is a disease associated with over-activation of complement and increased MAC deposition on retinal pigmented epithelial cells in the macula (McHarg et al. 2015).

Dysregulation of the terminal pathway has been implicated in the pathogenesis of a variety of autoimmune disorders as well as cancer. Paroxysmal nocturnal haemoglobinuria (PNH) is a rare and severe blood disease whereby patients critically lack CD59, the only membrane-bound regulator of MAC (Brodsky 2014). As a result, the red blood cells of these patients are vulnerable to complement attack and MAC-mediated haemolysis. Conversely, overexpression of CD59 has been implicated in the immune evasion mechanism in a variety of cancers, including breast, prostate, lung, ovarian, colorectal, and cervical carcinomas (Bjøge et al. 1994; Bjørge et al. 1997; Jarvis et al. 1997; Varsano et al. 1998; Jurianz et al. 1999; Gelderman et al. 2002; Fishelson et al. 2003; Donin et al. 2003). Overexpression of CD59 on cancer cells contributes to resistance of several monoclonal antibody anti-cancer therapies, including rituximab, a treatment for non-Hodgkin's B-cell lymphoma and chronic lymphocytic leukaemia (Treon et al. 2001; Smith 2003).

Targeting C5 Activation

Complement inhibition is a major focus of pharmaceutical research (Morgan and Harris 2015). Most therapeutic strategies focus on inhibiting the terminal pathway by targeting C5. Specifically blocking C5 reduces potential side-effects by preserving the immunoprotective and immunoregulatory functions of upstream complement cascades. This strategy is also versatile in that MAC is inhibited regardless of the upstream activating pathway. Furthermore, preventing C5 cleavage also suppresses the release of C5a, a potent anaphylatoxin implicated in inflammatory diseases including rheumatoid arthritis and sepsis (Woodruff et al. 2011).

Eculizumab is a potent anti-C5 monoclonal antibody therapeutic that prevents cleavage of C5 by the C5 convertase (Rother et al. 2007). Approved by the Food and Drug Association (FDA), eculizumab is used for the treatment of both PNH and aHUS (Hillmen et al. 2006; Zuber et al. 2012). The structure of C5 in complex with eculizumab revealed that the antibody specifically binds the MG7 domain of C5 and may act by sterically blocking the convertase binding-site (Fig. 7.6a) (Schatz-Jakobsen et al. 2016). Though effective for some, patients with genetic polymorphisms in the MG7 domain of C5 are not responsive to treatment (Nishimura et al. 2014). Eculizumab is also one of the most expensive therapies world-wide, costing nearly $500,000 per patient for one year of treatment (Jayasundara et al. 2019). With the prohibitive cost of therapy and some not responsive to treatment, there is clear need for alternative complement inhibitors that target the terminal pathway.

Structures of C5 in complex with virulence factors provide clues to designing new therapeutic alternatives that target convertase activity. Cobra venom factor (CVF) is a C3b homologue that binds Factor B in the presence of Mg^{2+} ions to mimic convertase activity (Vogel and Fritzinger 2010). Though there are currently no structures of the C5 convertase, the C5-CVF complex suggests that the MG4, MG5, and MG7 domains of C5 likely play a role in binding the endogenous convertase (Fig. 7.6b) (Laursen et al. 2011). While these data are in agreement with the eculizumab binding site, it is important to note that the C5-CVF complex remains only a proxy for investigating endogenous convertase interactions. Tick C5 inhibitors, such as OmCI from *Ornithodoros moubata* and the RaCI protein family from *Rhipicephalus appendiculatus*, block endogenous convertase-derived C5 activation, yet have no effect on CVF-based convertase activity (Fredslund et al. 2008; Jore et al. 2016). OmCI interacts with the C5 TED, CUB and C345C domains, far from both the eculizumab binding and convertase cleavage sites (Fig. 7.6c) (Jore et al. 2016). These structural data combined with competition assays suggest that C5 activation by endogenous convertases may involve an initial priming step that is not required for CVF-derived cleavage. Furthermore, high-resolution cryoEM studies of a novel class of tick inhibitors from the CirpT family reveal another way to block C5 activation by targeting C5 MG4 that could be exploited for drug discovery (Fig. 7.6d) (Reichhardt et al. 2020).

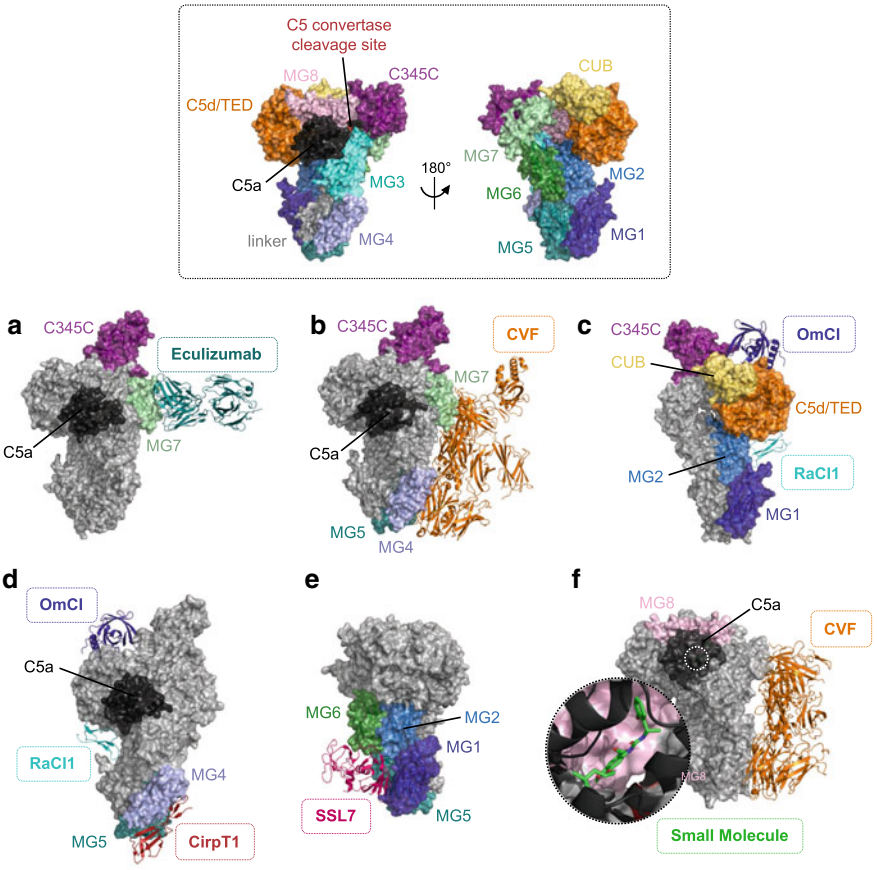


Fig. 7.6 Structural basis of blocking C5 activation. Surface representation of the C5 crystal structure (PDB 3CU7) coloured according to domains is shown above in a dotted box for reference. **a** Crystal structure of C5 (surface representation) in complex with eculizumab (dark teal ribbons) (PDB 5I5K). **b** Crystal structure of C5 (surface representation) in complex with cobra venom factor (CVF; orange ribbons) (PDB 3PVM). **c** Crystal structure of C5 (surface representation) in complex with tick inhibitors OmCI (dark blue ribbons) and RaCI1 (cyan ribbons) (PDB 5HCE). **d** Atomic model derived from the cryoEM structure of C5 (surface representation) in complex with tick inhibitors Cirp T1 (red ribbons), OmCI (dark blue ribbons), and RaCI1 (cyan ribbons) (PDB 6RQJ; EMD-4983). **e** Crystal structure of C5 (surface representation) in complex with Staphylococcal Superantigen-Like protein 7 (SSL7; magenta ribbons) (PDB 3KLS). **f** Atomic model derived from the cryoEM structure of C5 (surface representation) in complex with CVF (orange ribbons) and a small molecule that blocks cleavage of C5 (green sticks) (PDB 6I2X; EMD-4401). Circular inset shows a zoomed in view of the small molecule binding-site. For all panels C5a is in black, domains of C5 that form interaction interfaces are highlighted, and all other C5 domains are grey

Bacterial pathogens have also developed several strategies for regulating MAC to evade complement immune clearance. *Staphylococcus aureus* secretes Staphylococcal Superantigen-Like protein 7 (SSL7), which binds to C5 to block bactericidal activity from the MAC (Langley et al. 2005). A crystal structure of the C5-SSL7 complex revealed that SSL7 binds to C5 more than 70 Å away from the C5a cleavage site (Fig. 7.6e) (Laursen et al. 2010). Proteolytic cleavage of C5 into C5a and C5b may still occur in the presence of SSL7, however, subsequent binding of C6 to C5b to initiate MAC pore formation is inhibited.

Blocking Pore Formation

Once the terminal pathway is activated, there remains only one membrane-bound regulator of MAC: CD59. CD59 is present on all circulating cells and most human tissues, providing protection from autologous MAC attack (Meri et al. 1990, 1991). Comprised of three β-sheets and two α-helices, CD59 is a small 18–20 kDa extracellular glycoprotein (Fletcher et al. 1994; Leath et al. 2007). A complex network of five disulfide bridges maintains the protein's rigid, compact, and discoid structure (Sugita et al. 1993). Extending from the CD59 C-terminus, a glycosylphosphatidylinositol (GPI) lipid anchor tethers the molecule to the outer leaflet of the cell membrane. CD59 binds to the MACPF domains of either C9 or C8α to block subsequent incorporation of C9 molecules and to prevent membrane rupture (Meri et al. 1990; Ninomiya and Sims 1992; Farkas et al. 2002). Plasma-circulating forms of C8 and C9 do not bind CD59, suggesting that the binding sites are only exposed upon MAC formation (Huang et al. 2006). Although there are no structural data for CD59 inhibited MAC complexes, mutational studies have mapped the binding site to a hydrophobic cavity of CD59 (Bodian et al. 1997).

Similar to understanding C5 activation, the interaction of bacterial virulence factors with CD59 can inform new strategies for controlling complement activity. Secreted by over five genera of Gram-positive bacteria, cholesterol dependent cytolysins (CDCs) are pore-forming proteins that are structurally homologous to the MACPF domain of complement proteins (Fig. 7.3) (Tweten 2005; Hadders et al. 2007; Rosado et al. 2007). *Streptococcus intermedius* intermedilysin (ILY), *Gardenerella vaginalis* vaginolysin (VLY), and *Streptococcus mitis* lectinolysin (LLY) comprise a subset of these toxins that hijack CD59 to specifically target human cells (Lawrence et al. 2016). The interaction interface for ILY partially overlaps with the complement binding site on CD59 (Giddings et al. 2004). However, the primary interface lies within a β-hairpin of ILY that extends from domain 4 (D4), a domain that is absent in MAC proteins (Figs. 7.4 and 7.7) (Johnson et al. 2013). While the molecular basis for how CD59 binds C8α and C9 remains to be determined, CD59 may bind the newly extended β-hairpins of these MACPF-containing proteins to block membrane insertion.

Fig. 7.7 Crystal structure of the ILY-CD59 complex (PDB 4BIK). CD59 is in green; ILY is in blue with domain 4 (D4) highlighted in dark blue

Scavenging By-Products

In the absence of membranes, activated complement components can assemble off-target MAC precursor complexes. This 'soluble MAC' is referred to as sC5b9 and is used as a biomarker for systemic infections and acute meningococcal disease (Mook-Kanamori et al. 2014). Overactivation of complement overwhelms the C5 convertase and may impair its ability to orient MAC efficiently in target membranes (Heesterbeek et al. 2019). Consequently, MAC precursors are released into solution at different stages of assembly, which are then scavenged by blood-based chaperones clusterin and vitronectin to form stable complexes referred to as sC5b7, sC5b8 and sC5b9 (Preissner et al. 1989). Although primarily considered a soluble complex, it is possible that sC5b9 detected in vivo may also be attributed in part to membrane-bound material shed from damaged cells. A low resolution cryoEM structure of sC5b9 suggests that chaperones may act by shielding hydrophobic interactions of otherwise membrane-binding residues (Hadders et al. 2012). Atomic resolution details would

provide new insight into how regulators prevent the helix-to-hairpin transition of pore forming amino acids.

Future of Structure-Based Complement Therapeutics

High-resolution complexes of MAC and upstream regulators set the scene for future development of enhanced therapeutics. The potential for such therapies is highlighted by over a dozen anti-complement drugs currently in hundreds of active clinical trials across a range of drug types including recombinant proteins, peptides, small molecules, antibodies/biologics, and DNA (Morgan and Harris 2015; Mastellos et al. 2019). The potential applications are similarly diverse, ranging from overcoming the treatment failure with eculizumab in PNH and related diseases (Zipfel et al. 2019), to limiting damage from activated complement in bacterial infections (Mastellos et al. 2019; Koelman et al. 2019). Collectively the structures of apo and inhibited C5 provide a platform for development of highly specific complement inhibitors, which may either (i) selectively inhibit MAC pore formation without perturbing the C5a inflammatory signalling network, (ii) selectively impede an inflammation response from C5a, or (iii) halt the entire terminal complement system beyond C5 proteolysis. CryoEM has already started to play an important role in the development of C5 inhibitors, and the enhanced resolution achievable with modern techniques has enabled progress in small molecule drug discovery as well as in biologics (*e.g.* tick protein inhibitor) development. While the use of biologics suffers from substantial challenges of delivery, immunogenicity and pharmacokinetics, which are largely resolved by the use of small molecule drugs, it can be challenging to identify binding modes in large protein complexes by which small molecules could be effective.

A standout example of the small molecule approach is the recent identification of the first small molecule inhibitors of C5 complement protein by researchers at Novartis (Jendza et al. 2019). In this case, the binding site was initially identified by a careful analysis of structures and activities on a range of natural primate polymorphisms and engineered complement mutants. After failure to crystallise the complex for X-ray analysis, the remarkable binding mode of this compound series was only realised by the solution of a 3.35 Å complex with C5 and CVF by cryoEM (Fig. 7.6f). Designing compounds de novo that induce the highly buried conformation that this structure reveals remains challenging from a chemical biology perspective. This structure provides an important starting point for further in silico molecular modelling that may lead to the development of new and improved compounds. This study also highlights the value of cryoEM to understand the impact of polymorphisms by mapping disease-inducing or drug resistant mutations to key interfaces.

Given the complementary limitations of small molecule and protein drugs, constrained peptides provide an interesting middle ground. Peptide-derivative therapeutics limit immunogenicity while maintaining capacity to target large interaction sites including those with drug-resisting polymorphisms. Zilucoplan is a cyclic peptide discovered through mRNA display that allosterically and potently inhibits

C5 cleavage at a site orthogonal to that occupied by eculizumab (Ricardo et al. 2014). This therapeutic peptide has shown marked success in late stage clinical trials for myasthenia gravis in Asian populations, a key demographic of patients with genetic polymorphisms that cause eculizumab treatment failure (Howard et al. 2020). It is likely that cryoEM will play a key role in future structure-guided development of constrained peptide modulators of complement.

While most current therapeutic approaches target upstream regulators, targeting MAC formation itself has the potential to offer a more selective therapy that may be less susceptible to resistance from polymorphisms. With its ability to disentangle heterogeneous and flexible assemblies, cryoEM provides a powerful tool for resolving how CD59 blocks MAC. This information could prove transformational in designing small molecules or peptides which can enhance native CD59 activity, or discovery of agents which mimic CD59 MAC inhibition. A deeper understanding of the role of CD59 may also lead to drugs which can prevent the formation of bacterial lytic pores that hijack CD59 on human cells. Finally, selective activation of MAC on specific cell types is a plausible approach to enhance the efficacy of complement-mediated immunotherapies. Inhibitors of the CD59-MAC interaction may release the brake on complement-mediated lysis imposed by CD59 upregulation in treatment of cancers with antibody drugs such as rituximab. In particular, a more targeted therapy could restrict complement killing to relevant tissue types by conjugation to cell surface targeting motifs (Weiner 2010). To date these approaches have relied on biologics such as siRNA or immunogenic bacterial pore-forming toxin fragments (Geis et al. 2010; You et al. 2011). A cryoEM-enhanced understanding of how MAC is inhibited is likely to lead to knowledge-based design of novel advanced therapeutics that better tune complement activity.

Acknowledgements This work is supported by a CRUK Career Establishment Award (C26409/A16099) to D.B.; J.K.B. is supported by an EPSRC Doctoral Prize Fellowship and funded by an EPSRC Doctoral Training Program grant (EP/L015498/1); T.B.V. is funded by a BBSRC Doctoral Training Program grant (BB/M011178/1); E.W.T and D.B. are supported by a NIHR Imperial Biomedical Research Centre (BRC) Award (RDF01) and an EPSRC Impact Acceleration Account (EP/R511547/1). Figs were created with BioRender.com and rendered using UCSF Chimera (Pettersen et al. 2004) or PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

References

Aleshin AE, Schraufstatter IU, Stec B et al (2012) Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of Membrane Attack Complex (MAC). J Biol Chem 287:10210–10222. https://doi.org/10.1074/jbc.M111.327809

Bayly-Jones C, Bubeck D, Dunstone MA (2017) The mystery behind membrane insertion: a review of the complement membrane attack complex. Philos Trans R Soc B Biol Sci 372:20160221. https://doi.org/10.1098/rstb.2016.0221

Bexborn F, Andersson PO, Chen H et al (2008) The tick-over theory revisited: formation and regulation of the soluble alternative complement C3 convertase (C3(H2O)Bb). Mol Immunol 45:2370–2379. https://doi.org/10.1016/J.MOLIMM.2007.11.003

Bjøge L, Vedeler CA, Ulvestad E, Matre R (1994) Expression and function of CD59 on colonic adenocarcinoma cells. Eur J Immunol 24:1597–1603. https://doi.org/10.1002/eji.1830240722

Bjørge L, Hakulinen J, Wahlström T et al (1997) Complement-regulatory proteins in ovarian malignancies. Int J Cancer 70:14–25. https://doi.org/10.1002/(SICI)1097-0215(19970106)70:1%3c14::AID-IJC3%3e3.0.CO;2-9

Bodian DL, Davis SJ, Morgan BP, Rushmere NK (1997) Mutational analysis of the active site and antibody epitopes of the complement-inhibitory glycoprotein, CD59. J Exp Med 185:507–516. https://doi.org/10.1084/jem.185.3.507

Botto M, Kirschfink M, Macor P et al (2009) Complement in human diseases: lessons from complement deficiencies. Mol Immunol 46:2774–2783. https://doi.org/10.1016/J.MOLIMM.2009.04.029

Brannen CL, Sodetz JM (2007) Incorporation of human complement C8 into the membrane attack complex is mediated by a binding site located within the C8β MACPF domain. Mol Immunol 44:960–965. https://doi.org/10.1016/J.MOLIMM.2006.03.012

Brodsky RA (2014) Paroxysmal nocturnal hemoglobinuria. Blood 124:2804–2811. https://doi.org/10.1182/blood-2014-02-522128

Bubeck D (2014) The making of a macromolecular machine: assembly of the membrane attack complex. Biochemistry 53:1908–1915. https://doi.org/10.1021/bi500157z

Campbell AK, Daw RA, Hallett MB, Luzio JP (1981) Direct measurement of the increase in intracellular free calcium ion concentration in response to the action of complement. Biochem J 194:551–560. https://doi.org/10.1042/bj1940551

Carroll MC (2004) The complement system in regulation of adaptive immunity. Nat Immunol 5:981–986. https://doi.org/10.1038/ni1113

Cooper NR, Müller-Eberhard HJ (1970) The reaction mechanism of human C5 in immune hemolysis. J Exp Med 132:775–793. https://doi.org/10.1084/jem.132.4.775

DiScipio RG, Chakravarti DN, Muller-Eberhard HJ, Fey GH (1988) The structure of human complement component C7 and the C5b–7 complex. J Biol Chem 263:549–560

DiScipio RG, Smith CA, Muller-Eberhard HJ, Hugli TE (1983) The activation of human complement component C5 by a fluid phase C5 convertase. J Biol Chem 258:10629–10636

Donin N, Jurianz K, Ziporen L et al (2003) Complement resistance of human carcinoma cells depends on membrane regulatory proteins, protein kinases and sialic acid. Clin Exp Immunol 131:254–263. https://doi.org/10.1046/j.1365-2249.2003.02066.x

Dudkina NV, Spicer BA, Reboul CF et al (2016) Structure of the poly-C9 component of the complement membrane attack complex. Nat Commun 7:10588. https://doi.org/10.1038/ncomms10588

Dunkelberger JR, Song W-C (2010) Complement and its role in innate and adaptive immune responses. Cell Res 20:34–50. https://doi.org/10.1038/cr.2009.139

Fang C, Manes TD, Liu L et al (2019) ZFYVE21 is a complement-induced Rab5 effector that activates non-canonical NF-κB via phosphoinosotide remodeling of endosomes. Nat Commun 10:2247. https://doi.org/10.1038/s41467-019-10041-2

Farkas I, Baranyi L, Ishikawa Y et al (2002) CD59 blocks not only the insertion of C9 into MAC but inhibits ion channel formation by homologous C5b–8 as well as C5b–9. J Physiol 539:537–545. https://doi.org/10.1113/jphysiol.2001.013381

Faruqi A, Henderson R (2007) Electronic detectors for electron microscopy. Curr Opin Struct Biol 17:549–555. https://doi.org/10.1016/J.SBI.2007.08.014

Fishelson Z, Donin N, Zell S et al (2003) Obstacles to cancer immunotherapy: expression of membrane complement regulatory proteins (mCRPs) in tumors. Mol Immunol 40:109–123. https://doi.org/10.1016/S0161-5890(03)00112-3

Fletcher CM, Harrison RA, Lachmann PJ, Neuhaus D (1994) Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. Structure 2:185–199. https://doi.org/10.1016/S0969-2126(00)00020-4

Fredslund F, Laursen NS, Roversi P et al (2008) Structure of and influence of a tick complement inhibitor on human complement component 5. Nat Immunol 9:753–760. https://doi.org/10.1038/ni.1625

Geis N, Zell S, Rutz R et al (2010) Inhibition of membrane complement inhibitor expression (CD46, CD55, CD59) by siRNA sensitizes tumor cells to complement attack in vitro. Curr Cancer Drug Targets 10:922–931. https://doi.org/10.2174/156800910793357952

Gelderman KA, Blok VT, Fleuren GJ, Gorter A (2002) The inhibitory effect of CD46, CD55, and CD59 on complement activation after immunotherapeutic treatment of cervical carcinoma cells with monoclonal antibodies or bispecific monoclonal antibodies. Lab Investig 82:483–493. https://doi.org/10.1038/labinvest.3780441

Gerard NP, Gerard C (1991) The chemotactic receptor for human C5a anaphylatoxin. Nature 349:614–617. https://doi.org/10.1038/349614a0

Giddings KSK, Zhao J, Sims PPJ, Tweten RKR (2004) Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. Nat Struct Mol Biol 11:1173–1178. https://doi.org/10.1038/nsmb862

Guo R-F, Ward PA (2005) Role of C5a in inflammatory responses. Annu Rev Immunol 23:821–852. https://doi.org/10.1146/annurev.immunol.23.021704.115835

Hadders M, a, Beringer DX, Gros P, (2007) Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. Science 317:1552–1554. https://doi.org/10.1126/science.1147103

Hadders MA, Bubeck D, Roversi P et al (2012) Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. Cell Rep 1:200–207. https://doi.org/10.1016/j.celrep.2012.02.003

Harriman GR, Esser AF, Podack ER et al (1981) The role of C9 in complement-mediated killing of Neisseria. J Immunol 127:2386–2390

Heesterbeek DA, Bardoel BW, Parsons ES et al (2019) Bacterial killing by complement requires membrane attack complex formation via surface-bound C5 convertases. EMBO J 38:e99852. https://doi.org/10.15252/embj.201899852

Hillmen P, Young NS, Schubert J et al (2006) The Complement Inhibitor Eculizumab in paroxysmal nocturnal Hemoglobinuria. N Engl J Med 355:1233–1243. https://doi.org/10.1056/NEJMoa061648

Hoover DL, Berger M, Nacy CA et al (1984) Killing of Leishmania tropica amastigotes by factors in normal human serum. J Immunol 132:893–897

Howard JF, Nowak RJ, Wolfe GI et al (2020) Clinical effects of the self-administered subcutaneous complement inhibitor Zilucoplan in patients with moderate to severe generalized Myasthenia Gravis. JAMA Neurol 17:e195125. https://doi.org/10.1001/jamaneurol.2019.5125

Huang Y, Qiao F, Abagyan R et al (2006) Defining the CD59-C9 binding interaction. J Biol Chem 281:27398–27404. https://doi.org/10.1074/jbc.M603690200

Jane-wit D, Surovtseva YV, Qin L et al (2015) Complement membrane attack complexes activate noncanonical NF-κB by forming an Akt+ NIK+ signalosome on Rab5+ endosomes. Proc Natl Acad Sci 112:9686–9691. https://doi.org/10.1073/pnas.1503535112

Jarvis GA, Li J, Hakulinen J et al (1997) Expression and function of the complement membrane attack complex inhibitor protectin (CD59) in human prostate cancer. Int J Cancer 71:1049–1055. https://doi.org/10.1002/(SICI)1097-0215(19970611)71:6%3c1049::AID-IJC22%3e3.0.CO;2-7

Jayasundara K, Hollis A, Krahn M et al (2019) Estimating the clinical cost of drug development for orphan versus non-orphan drugs. Orphanet J Rare Dis 14:12. https://doi.org/10.1186/s13023-018-0990-4

Jendza K, Kato M, Salcius M et al (2019) (2019) A small-molecule inhibitor of C5 complement protein. Nat Chem Biol 157(15):666–668. https://doi.org/10.1038/s41589-019-0303-9

Johnson S, Brooks NJ, Smith RAG et al (2013) Structural basis for recognition of the pore-forming toxin intermedilysin by human complement receptor CD59. Cell Rep 3:1369–1377. https://doi.org/10.1016/j.celrep.2013.04.029

Jore MM, Johnson S, Sheppard D et al (2016) Structural basis for therapeutic inhibition of complement C5. Nat Struct Mol Biol 23:378–386. https://doi.org/10.1038/nsmb.3196

Jurianz K, Maslak S, Garcia-Schüler H et al (1999) Neutralization of complement regulatory proteins augments lysis of breast carcinoma cells targeted with rhumAb anti-HER2. Immunopharmacology 42:209–218. https://doi.org/10.1016/S0162-3109(99)00006-5

Koelman DLH, Brouwer MC, van de Beek D (2019) Targeting the complement system in bacterial meningitis. Brain 142:3325–3337. https://doi.org/10.1093/brain/awz222

Konovalova A, Kahne DE, Silhavy TJ (2017) Outer membrane biogenesis. Annu Rev Microbiol 71:539–556. https://doi.org/10.1146/annurev-micro-090816-093754

Langley R, Wines B, Willoughby N et al (2005) The Staphylococcal Superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-FcαRI binding and serum killing of bacteria. J Immunol 174:2926–2933. https://doi.org/10.4049/JIMMUNOL.174.5.2926

Laursen NS, Andersen KR, Braren I et al (2011) Substrate recognition by complement convertases revealed in the C5-cobra venom factor complex. EMBO J 30:606–616. https://doi.org/10.1038/emboj.2010.341

Laursen NS, Gordon N, Hermans S et al (2010) Structural basis for inhibition of complement C5 by the SSL7 protein from Staphylococcus aureus. Proc Natl Acad Sci 107:3681–3686. https://doi.org/10.1073/PNAS.0910565107

Law RHP, Lukoyanova N, Voskoboinik I et al (2010) The structural basis for membrane binding and pore formation by lymphocyte perforin. Nature 468:447–451. https://doi.org/10.1038/nature09518

Lawrence SL, Gorman MA, Feil SC et al (2016) Structural basis for receptor recognition by the human CD59-responsive cholesterol-dependent cytolysins. Structure 24:1488–1498. https://doi.org/10.1016/j.str.2016.06.017

Leath KJ, Johnson S, Roversi P et al (2007) High-resolution structures of bacterially expressed soluble human CD59. Acta Crystallogr Sect F 63:648–652. https://doi.org/10.1107/S1744309107033477

Li X, Mooney P, Zheng S et al (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods 10:584–590. https://doi.org/10.1038/nmeth.2472

Lovelace LL, Cooper CL, Sodetz JM, Lebioda L (2011) Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. J Biol Chem 286:17585–17592. https://doi.org/10.1074/jbc.M111.219766

Lueck K, Wasmuth S, Williams J et al (2011) Sub-lytic C5b–9 induces functional changes in retinal pigment epithelial cells consistent with age-related macular degeneration. Eye 25:1074–1082. https://doi.org/10.1038/eye.2011.109

Mastellos DC, Ricklin D, Lambris JD (2019) Clinical promise of next-generation complement therapeutics. Nat Rev Drug Discov 18:707–729. https://doi.org/10.1038/s41573-019-0031-6

McHarg S, Clark SJ, Day AJ, Bishop PN (2015) Age-related macular degeneration and the role of the complement system. Mol Immunol 67:43–50. https://doi.org/10.1016/J.MOLIMM.2015.02.032

Menny A, Serna M, Boyd CM et al (2018) CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. Nat Commun 9:5316. https://doi.org/10.1038/s41467-018-07653-5

Meri S, Morgan BP, Davies A et al (1990) Human protectin (CD59), an 18000–20000 MW complement lysis restricting factor, inhibits C5b–8 catalysed insertion of C9 into lipid bilayers. Immunology 71:1–9

Meri S, Waldmann H, Lachmann PJ (1991) Distribution of protectin (CD59), a complement membrane attack inhibitor, in normal human tissues. Lab Investig 65:532–537

Mook-Kanamori BB, Brouwer MC, Geldhoff M et al (2014) Cerebrospinal fluid complement activation in patients with pneumococcal and meningococcal meningitis. J Infect 68:542–547. https://doi.org/10.1016/J.JINF.2013.12.016

Morgan BP (1989) Complement membrane attack on nucleated cells: resistance, recovery and non-lethal effects. Biochem J 264:1–14. https://doi.org/10.1042/BJ2640001

Morgan BP (2016) The membrane attack complex as an inflammatory trigger. Immunobiology 221:747–751. https://doi.org/10.1016/J.IMBIO.2015.04.006

Morgan BP, Harris CL (2015) Complement, a target for therapy in inflammatory and degenerative diseases. Nat Rev Drug Discov 14:857–877. https://doi.org/10.1038/nrd4657

Nakamura M, Okada H, Sasaki H et al (1996) Quantification of the CD55 and CD59, membrane inhibitors of complement on HIV-1 particles as a function of complement-mediated Virolysis. Microbiol Immunol 40:561–567. https://doi.org/10.1111/j.1348-0421.1996.tb01109.x

Ninomiya H, Sims PJ (1992) The human complement regulatory protein CD59 binds to the alpha-chain of C8 and to the 'b'domain of C9. J Biol Chem 267:13675–13680

Nishimura J-I, Yamamoto M, Hayashi S et al (2014) Genetic variants in C5 and poor response to Eculizumab. N Engl J Med 370:632–641. https://doi.org/10.1056/NEJMoa1311084

Noris M, Remuzzi G (2009) Atypical Hemolytic-uremic syndrome. N Engl J Med 361:1676–1687. https://doi.org/10.1056/NEJMra0902814

Pangburn MK, Schreiber RD, Müller-Eberhard HJ (1981) Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. J Exp Med 154:856–867. https://doi.org/10.1084/jem.154.3.856

Parker CL, Sodetz JM (2002) Role of the human C8 subunits in complement-mediated bacterial killing: evidence that C8γ is not essential. Mol Immunol 39:453–458. https://doi.org/10.1016/S0161-5890(02)00121-9

Parsons ES, Stanley GJ, Pyne ALB et al (2019) Single-molecule kinetics of pore assembly by the membrane attack complex. Nat Commun 10:2066. https://doi.org/10.1038/s41467-019-10058-7

Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/JCC.20084

Podack E, Tschoop J, Muller-Eberhard H (1982) Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b–8 assembly. J Exp Med 156:268–282. https://doi.org/10.1084/jem.156.1.268

Podack ER (1984) Molecular composition of the tubular structure of the membrane attack complex of complement. J Biol Chem 259:8641–8647

Podack ER, Esser AF, Biesecker G, Müller-Eberhard HJ (1980) Membrane attack complex of complement: a structural analysis of its assembly. J Exp Med 151:301–313. https://doi.org/10.1084/jem.151.2.301

Polekhina G, Giddings KS, Tweten RK, Parker MW (2005) Insights into the action of the superfamily of cholesterol-dependent cytolysins from studies of intermedilysin. Proc Natl Acad Sci 102:600–605. https://doi.org/10.1073/pnas.0403229101

Preissner KP, Podack ER, Müller-Eberhard HJ (1989) SC5b-7, SC5b-8 and SC5b-9 complexes of complement: ultrastructure and localization of the S-protein (vitronectin) within the macromolecules. Eur J Immunol 19:69–75. https://doi.org/10.1002/eji.1830190112

Preissner KT, Podack ER, Müller-Eberhard HJ (1985) The membrane attack complex of complement: relation of C7 to the metastable membrane binding site of the intermediate complex C5b–7. J Immunol 135:445–451

Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290–296. https://doi.org/10.1038/nmeth.4169

Reichhardt MP, Johnson S, Tang T et al (2020) An inhibitor of complement C5 provides structural insights into activation. Proc Natl Acad Sci 117:362–370. https://doi.org/10.1073/PNAS.1909973116

Ricardo A, Arata M, DeMarco SJ et al (2014) Development of RA101348, a potent cyclic peptide inhibitor of C5 for complement-mediated diseases. Blood 124:2936–2936. https://doi.org/10.1182/blood.V124.21.2936.2936

Ricklin D, Hajishengallis G, Yang K, Lambris JD (2010) Complement: a key system for immune surveillance and homeostasis. Nat Immunol 11:785–797. https://doi.org/10.1038/ni.1923

Ricklin D, Reis ES, Lambris JD (2016) Complement in disease: a defence system turning offensive. Nat Rev Nephrol 12:383–401. https://doi.org/10.1038/nrneph.2016.70

Rosado CJ, Buckle AM, Law RHP et al (2007) A Common fold mediates vertebrate defense and bacterial attack. Science 317:1548–1551. https://doi.org/10.1126/science.1144706

Rossjohn J, Feil SC, McKinstry WJ et al (1997) Structure of a cholesterol-binding, Thiol-activated Cytolysin and a model of its membrane form. Cell 89:685–692. https://doi.org/10.1016/S0092-8674(00)80251-2

Rother RP, Rollins SA, Mojcik CF et al (2007) Discovery and development of the complement inhibitor eculizumab for the treatment of paroxysmal nocturnal hemoglobinuria. Nat Biotechnol 25:1256–1264. https://doi.org/10.1038/nbt1344

Roumenina LT, Loirat C, Dragon-Durey M-A et al (2011) Alternative complement pathway assessment in patients with atypical HUS. J Immunol Methods 365:8–26. https://doi.org/10.1016/J.JIM.2010.12.020

Schatz-Jakobsen JA, Zhang Y, Johnson K et al (2016) Structural basis for Eculizumab-mediated inhibition of the complement terminal pathway. J Immunol 197:337–344. https://doi.org/10.4049/JIMMUNOL.1600280

Schneider MC, Exley RM, Ram S et al (2007) Interactions between Neisseria meningitidis and the complement system. Trends Microbiol 15:233–240. https://doi.org/10.1016/j.tim.2007.03.005

Serna M, Giles JL, Morgan BP, Bubeck D (2016) Structural basis of complement membrane attack complex formation. Nat Commun 7:10587. https://doi.org/10.1038/ncomms10587

Sharp TH, Koster AJ, Gros P (2016) Heterogeneous MAC initiator and pore structures in a lipid bilayer by phase-plate Cryo-electron tomography. Cell Rep 15:1–8. https://doi.org/10.1016/j.celrep.2016.03.002

Shatursky O, Heuck AP, Shepard LA et al (1999) The mechanism of membrane insertion for a cholesterol-dependent cytolysin: a novel paradigm for pore-forming toxins. Cell 99:293–299. https://doi.org/10.1016/s0092-8674(00)81660-8

Shepard LA, Heuck AP, Hamman BD et al (1998) Identification of a membrane-spanning domain of the Thiol-activated Pore-forming Toxin Clostridium perfringens Perfringolysin O: an α-Helical to β-sheet transition identified by fluorescence spectroscopy. Biochemistry 37:14563–14574. https://doi.org/10.1021/BI981452F

Silversmith RE, Nelsestuen GL (1986) Interaction of complement proteins C5b–6 and C5b–7 with phospholipid vesicles: effects of phospholipid structural features. Biochemistry 25:7717–7725. https://doi.org/10.1021/bi00371a065

Smith MR (2003) Rituximab (monoclonal anti-CD20 antibody): mechanisms of action and resistance. Oncogene 22:7359–7368. https://doi.org/10.1038/sj.onc.1206939

Spicer BA, Law RHP, Caradoc-Davies TT et al (2018) The first transmembrane region of complement component-9 acts as a brake on its self-assembly. Nat Commun 9:3266. https://doi.org/10.1038/s41467-018-05717-0

Steckel EW, Welbaum BE, Sodetz JM (1983) Evidence of direct insertion of terminal complement proteins into cell membrane bilayers during cytolysis. Labeling by a photosensitive membrane probe reveals a major role for the eighth and ninth components. J Biol Chem 258:4318–4324

Stewart JL, Kolb WP, Sodetz JM (1987) Evidence that C5b recognizes and mediates C8 incorporation into the cytolytic complex of complement. J Immunol 139:1960–1964

Sugita Y, Nakano Y, Oda E et al (1993) Determination of carboxyl-terminal residue and bisulfide bonds of MACIF (CD59), a Glycosyl-phosphatidylinositol-anchored membrane protein. J Biochem 114:473–477. https://doi.org/10.1093/oxfordjournals.jbchem.a124202

Tomlinson S, Taylor PW, Morgan BP, Luzio JP (1989) Killing of gram-negative bacteria by complement. Fractionation of cell membranes after complement C5b–9 deposition on to the surface of Salmonella minnesota Re595. Biochem J 263:505–511. https://doi.org/10.1042/bj2630505

Treon SP, Mitsiades C, Mitsiades N et al (2001) Tumor cell expression of CD59 is associated with resistance to CD20 serotherapy in patients with B-cell malignancies. J Immunother 24:263–271

Triantafilou K, Hughes TR, Triantafilou M, Morgan BP (2013) The complement membrane attack complex triggers intracellular Ca2+ fluxes leading to NLRP3 inflammasome activation. J Cell Sci 126:2903–2913. https://doi.org/10.1242/JCS.124388

Tschopp J, Podack ER, Müller-Eberhard HJ (1985) The membrane attack complex of complement: C5b–8 complex as accelerator of C9 polymerization. J Immunol 134:495–499

Tweten RK (2005) Cholesterol-dependent Cytolysins, a family of versatile pore-forming toxins. Infect Immun 73:6199–6209. https://doi.org/10.1128/IAI.73.10.6199-6209.2005

Varsano R, Shapiro et al (1998) Human lung cancer cell lines express cell membrane complement inhibitory proteins and are extremely resistant to complement-mediated lysis; a comparison with normal human respiratory epithelium in vitro , and an insight into mechanism(s) of resistance. Clin Exp Immunol 113:173–182. https://doi.org/10.1046/j.1365-2249.1998.00581.x

Vogel C-W, Fritzinger DC (2010) Cobra venom factor: structure, function, and humanization for therapeutic complement depletion. Toxicon 56:1198–1222. https://doi.org/10.1016/J.TOXICON.2010.04.007

Weiner GJ (2010) Rituximab: mechanism of action. Semin Hematol 47:115–123. https://doi.org/10.1053/J.SEMINHEMATOL.2010.01.011

Woodruff TM, Nandakumar KS, Tedesco F (2011) Inhibiting the C5–C5a receptor axis. Mol Immunol 48:1631–1642. https://doi.org/10.1016/J.MOLIMM.2011.04.014

Wu S, Armache J-P, Cheng Y (2016) Single-particle cryo-EM data acquisition by using direct electron detection camera. Microscopy 65:35–41. https://doi.org/10.1093/jmicro/dfv355

Xie CB, Jane-wit D, Pober JS (2020) Complement membrane attack complex: new roles, mechanisms of action, and therapeutic targets. Am J Pathol. https://doi.org/10.1016/J.AJPATH.2020.02.006

Yorulmaz S, Tabaei SR, Kim M et al (2015) Membrane attack complex formation on a supported lipid bilayer: initial steps towards a CARPA predictor nanodevice. Eur J Nanomedicine 7:245–255. https://doi.org/10.1515/ejnm-2015-0016

You T, Hu W, Ge X et al (2011) Application of a novel inhibitor of human CD59 for the enhancement of complement-dependent cytolysis on cancer cells. Cell Mol Immunol 8:157–163. https://doi.org/10.1038/cmi.2010.35

Zipfel PF, Wiech T, Rudnick R et al (2019) Complement inhibitors in clinical trials for glomerular diseases. Front Immunol 10:2166. https://doi.org/10.3389/fimmu.2019.02166

Zivanov J, Nakane T, Forsberg BO et al (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7:e42166. https://doi.org/10.7554/eLife.42166

Zuber J, Fakhouri F, Roumenina LT et al (2012) Use of eculizumab for atypical haemolytic uraemic syndrome and C3 glomerulopathies. Nat Rev Nephrol 8:643–657. https://doi.org/10.1038/nrneph.2012.214

Chapter 8
Architecture and Assembly of the Bacterial Flagellar Motor Complex

Yusuke V. Morimoto and Tohru Minamino

Abstract One of the central systems responsible for bacterial motility is the flagellum. The bacterial flagellum is a macromolecular protein complex that is more than five times the cell length. Flagella-driven motility is coordinated via a chemosensory signal transduction pathway, and so bacterial cells sense changes in the environment and migrate towards more desirable locations. The flagellum of *Salmonella enterica* serovar Typhimurium is composed of a bi-directional rotary motor, a universal joint and a helical propeller. The flagellar motor, which structurally resembles an artificial motor, is embedded within the cell envelop and spins at several hundred revolutions per second. In contrast to an artificial motor, the energy utilized for high-speed flagellar motor rotation is the inward-directed proton flow through a transmembrane proton channel of the stator unit of the flagellar motor. The flagellar motor realizes efficient chemotaxis while performing high-speed movement by an ingenious directional switching mechanism of the motor rotation. To build the universal joint and helical propeller structures outside the cell body, the flagellar motor contains its own protein transporter called a type III protein export apparatus. In this chapter we summarize the structure and assembly of the *Salmonella* flagellar motor complex.

Keywords Bacterial flagellum · Flagellar assembly · Chemotaxis · Stator · Rotor · Torque generation · Type III protein export apparatus · Proton motive force

Y. V. Morimoto
Department of Physics and Information Technology, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
e-mail: yvm001@bio.kyutech.ac.jp

T. Minamino (✉)
Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: tohru@fbs.osaka-u.ac.jp

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_8

Introduction

The bacterial flagellum, which is a large filamentous assembly, is one of the organelles involved in motility in various environments such as liquids and solid surfaces. *Salmonella enterica* serovar Typhimurium (*Salmonella*) has several flagella on the cell body. Each flagellum is a supramolecular protein complex made of about 30 different proteins, and the copy number of each component protein varies from a few to tens of thousands. More than 60 proteins are involved in flagellar construction, force generation and chemotactic behavior in *Salmonella* (Table 8.1). The *Salmonella* flagellum is made of five distinct structural and functional parts: a base body, a hook, a junction, a filament and a filament cap (Fig. 8.1). The base body is embedded in the cellular membranes and works as a bi-directional rotary motor energized by the electrochemical potential difference of protons (H^+) across the cytoplasmic membrane (proton motive force, PMF). The hook, junction, filament and filament cap structures are located outside the bacterial cell body. The filament acts as a helical propeller to generate propulsion. The hook functions as a universal joint that smoothly transmits rotational force produced by the bi-directional rotary motor to the helical propeller. The junction connects the hook and filament. The filament cap supports the self-assembly of the filament protein, flagellin, into the long helical filament structure (Minamino and Imada 2015; Nakamura and Minamino 2019).

The *Salmonella* flagellar motor rotates in both counterclockwise (CCW) and clockwise (CW) directions. When all the flagellar motors rotate in CCW direction, long helical flagellar filaments form a bundle structure behind a cell body, due to bending flexibility of each hook structure. As a result, the flagellar bundle produces thrust enough to allow *Salmonella* cells to go straight. When one or more motors switch the rotational direction from CCW to CW, the bundle structure is partially disrupted, and so the cells stop swimming and change the swimming direction (Fig. 8.2a) (Morimoto and Minamino 2014). Each flagellar motor is placed under a control of the sensory signaling network, which modulates the switching frequency of flagellar motor rotation, and hence bacterial cells carry out chemotaxis by a biased random walk toward various chemicals, pH and temperature to move toward more suitable stimuli (Fig. 8.2b, c) (Berg 2003).

Methyl-accepting chemotaxis protein (MCP) is a transmembrane protein with a large cytoplasmic domain and senses temporal changes in environmental stimuli. The adapter protein CheW binds to the cytoplasmic domain of the MCP, allowing the CheA kinase to efficiently associate with the MCP. The cytoplasmic domain of MCP controls the autophosphorylation activity of the CheA kinase to generate the chemotaxis signal. A phosphorylated form of the CheA kinase transfers its phosphate group to CheY, which acts as the chemotaxis signaling protein. Phosphorylated CheY (CheY-P) binds to two cytoplasmic rotor component proteins, FliM and FliN, in the flagellar motor. As a result, the motor switches the direction of rotation from CCW to CW. The CheZ phosphatase accelerates the dephosphorylation of CheY-P to induce the dissociation of CheY from the motor to allow the motor to rotate CCW again. This chemotaxis signaling network adapts to changes in the concentration of

Table 8.1 Bacterial flagellar proteins in *Salmonella*

Protein	Function	Size (kD)
FliF	MS ring	61
FliG	C ring, torque generation, directional switch	37
FliM	C ring, directional switch	38
FliN	C ring, directional switch	15
FlhA	Type III export gate protein, energy transducer, protein export switch	75
FlhB	Type III export gate protein, protein export switch	42
FliH	Type III export apparatus, peripheral stalk	26
FliI	Type III export apparatus, ATPase	49
FliJ	Type III export apparatus, central stalk	17
FliO	Scaffolding protein for the assembly of the $FliP_5FliR_1$ complex	13
FliP	Type III export gate protein	27
FliQ	Type III export gate protein	10
FliR	Type III export gate protein	29
FlhE	Plug for a proton channel in a type III export apparatus	14
FlgN	Export chaperone specific for FlgK and FlgL	16
FliS	Export chaperone specific for FliC	15
FliT	Export chaperone specific for FliD	14
FliE	Basal body protein connecting the MS ring and the proximal rod	11
FlgB	Proximal rod	15
FlgC	Proximal rod	14
FlgF	Proximal rod	26
FlgG	Distal rod	28
FlgJ	Rod cap, Muramidase	34
FlgI	P ring	38
FlgA	Periplasmic chaperone for P ring assembly	24
FlgH	L ring	25
FlgD	Hook cap	24
FlgE	Hook	42
FliK	Hook-length control	42
FlgK	Hook-filament junction	59
FlgL	Hook-filament junction	34
FliD	Filament cap	50
FliC	Filament (H1 flagellin)	52
FljB	Filament (H2 flagellin)	53
FljA	Negative regulator of FliC expression	20
MotA	Stator, transmembrane proton channel	32

(continued)

Table 8.1 (continued)

Protein	Function	Size (kD)
MotB	Stator, transmembrane proton channel	34
FliL	Stator associated protein	17
FlhC	Master regulator for flagellar genes	22
FlhD	Master regulator for flagellar genes	13
FliA	Sigma factor/Chaperone specific for FlgM	27
FliZ	Positive regulator for flagellar gene expression	22
FlgM	Anti-sigma factor	11
CheZ	Chemotaxis protein	24
CheY	Chemotaxis protein	14
CheB	Chemotaxis protein	38
CheR	Chemotaxis protein	33
CheW	Chemotaxis protein	18
CheA	Chemotaxis protein	73
Tsr	Methyl-accepting chemotaxis protein	60
Tar	Methyl-accepting chemotaxis protein	60
Aer	Methyl-accepting chemotaxis protein	55

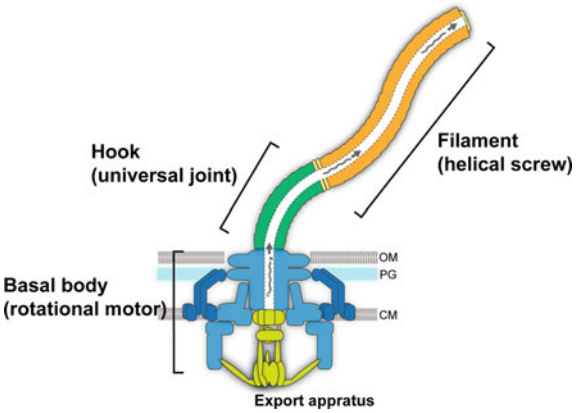


Fig. 8.1 Schematic diagram of the bacterial flagellar motor complex structure. The flagellum is composed of a filament, a hook and a basal body. Each building block synthesized in the cytoplasm is translocated via a type III protein export apparatus into the central channel of the growing axial structure. OM—outer membrane, PG—peptidoglycan layer, CM—cytoplasmic membrane

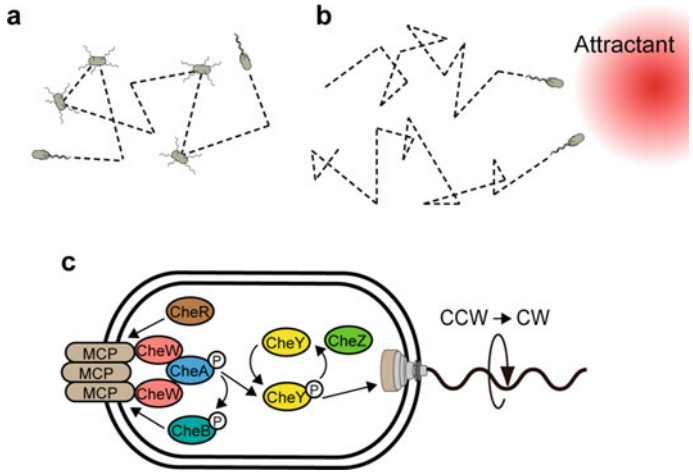


Fig. 8.2 **a** Swimming pattern of *Salmonella* cells. *Salmonella* cells undergoes continuous run-and-stop pattern, thereby showing random walk motions in liquids. When the flagellar motor rotates in counterclockwise (CCW) direction, *Salmonella* cells can go straight. In contrast, when the motor switches its rotational direction from CCW to counterclockwise (CW), the cells stop and change the swimming direction. **b** Biased random walk motion to attractants. *Salmonella* cells sense gradients of chemical attractants in the environments to move towards more favorable stimuli. **c** Schematic diagram of a sensory signaling transduction pathway responsible for chemotaxis. The *Salmonella* phosphorelay signal transduction network is formed by methyl-accepting chemotaxis protein (MCP), the CheW adopter protein, the CheA kinase, the CheY chemotaxis signaling protein, the CheZ phosphatase, the CheB methylesterase and the CheR methyltransferase. CheY-P binds to the switch complex of the flagellar motor to chnage the direction of flagellar motor rotation from CCW to CW

chemical stimuli over a wide dynamic range. CheB and CheR are involved in the adaption mechanism. CheR is the methyltransferase to induce the methylation of the cytoplasmic domain of the MCP to modulate the autophosphorylation activity of the CheA kinase. The CheA kinase also transfers the phosphate group to the CheB methylesterase to promote the demethylation of the MCP to reduce the probability of autophosphorylation of the CheA kinase (Fig. 8.2c) (Bi and Sourjik 2018).

The *Salmonella* flagellar motor consists of a rotor and about ten stator units placed around the rotor. The rotor is a macromolecular ring complex made of four flagellar proteins, FliF, FliG, FliM and FliN. The transmembrane MS ring is formed by FliF. FliG, FliM and FliN assemble into the C ring on the cytoplasmic surface of the MS ring in this order. The C ring also works as a directional switching device to switch between CCW and CW states of the flagellar motor. Each stator unit is composed of two integral membrane proteins, MotA and MotB, which form a proton channel complex to convert the energy of the proton flow through the channel into the rotational force (Minamino et al. 2018; Nakamura and Minamino 2019).

To construct the axial structure of the flagellum such as the hook and filament outside the cell body, the basal body has a type III protein export apparatus at its proximal end (Macnab 2003). The type III protein export apparatus is also a macromolecular protein complex consisting of a transmembrane export gate complex powered by PMF across the cytoplasmic membrane and an associated cytoplasmic ATPase ring complex and transports building blocks of the axial structure from the cytoplasm to the distal end of the nascent structure where each building block self-assembles into the axial structure. The export gate complex is located inside the central pore of the MS ring whereas the cytoplasmic ATPase ring complex associates with the C ring (Minamino 2014, 2018). In this chapter, we describe the structure and function of the bacterial flagellar motor complex in *Salmonella.*

Axial Structure

The *Salmonella* flagellum is composed of several basal body rings and an axial structure (Fig. 8.1). The axial structure is a helical, tubular structure consisting of six distinct structural parts: the proximal rod (FlgB, FlgC, FlgF), the distal rod (FlgG), the hook (FlgE), the junction (FlgK, FlgL), the filament (FliC, FljB), and the filament cap (FliD). The N- and C-terminal regions of each building block are disordered in solution but adopt a coiled coil in the inner core domain (D0) of the building block when they assemble into the axial structure (Fig. 8.3a) (Furukawa et al. 2002; Saijo-Hamano et al. 2004; Samatey et al. 2001). Hydrophobic interactions between the D0 domains of the building blocks are essential not only for the self-assembly of the axial structure but also for structural and mechanical stability of the entire tubular structure (Yonekura et al. 2003; Fujii et al. 2009; Maki-Yonekura et al. 2010; Saijo-Hamano et al. 2019). Except for the cap structure, which has five-fold rotational symmetry (Yonekura et al. 2000), the axial structure is a helical assembly composed of 11 protofilaments (Fig. 8.3b) (Yonekura et al. 2003; Fujii et al. 2009, 2017; Maki-Yonekura et al. 2010; Kato et al. 2019). The basic helical line that passes through all the subunits is called a 1-start helix, and there are approximately 11 subunits per two turns of the 1-start helix (Fig. 8.3b).

The rod is straight and rigid and acts as a drive shaft of the flagellar motor (Fujii et al. 2017). The hook is supercoiled and highly flexible in bending and functions as a universal joint (Samatey et al. 2004; Fujii et al. 2009; Kato et al. 2019). The filament is also supercoiled but rigid against bending and so functions as a helical propeller (Yonekura et al. 2003; Maki-Yonekura et al. 2010). The *Salmonella* filament normally adopts a left-handed supercoil to form a flagellar bundle structure behind a cell body for straight swimming and switches from the left-handed to right-handed supercoils when the flagellar motor switches the direction of rotation from CCW to CW (Calladine 1975, 1976; Maki-Yonekura et al. 2010). Since the rod and hook structurally look similar to each other, the hook is directly connected with the rod (Chevance et al. 2007; Fujii et al. 2017). In contrast, because there are structural and mechanical differences between the hook and filament, the junction is a cushioning

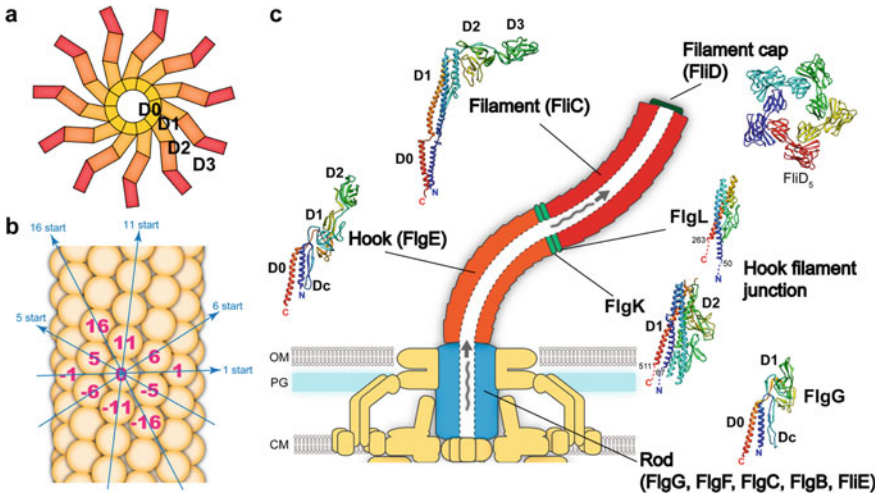


Fig. 8.3 Axial structure in the bacterial flagellar motor complex. **a** Schematic diagram of cross section of the flagellar filament. Flagellar filament protein is composed of domains D0, D1, D2 and D3. **b** Arrangement of building blocks in the flagellar axial structure composed of 11 protofilaments. Arrows show the helical lines. The 11-start helical line constitutes each protofilament. The numbers of the subunit starting from the subunit 0 along the 1-start helical line are shown. **c** Representative atomic structures of flagellar axial component proteins, FlgG (PDB ID, 6JZR), FlgE (PDB ID, 6KFK), FlgK (PDB ID, 2D4Y), FlgL (PDB ID, 2D4X), FliC (PDB ID, 1UCU) and FliD (PDB ID, 5H5T) derived from *Salmonella.* CM—cytoplasmic membrane, PG—peptidoglycan layer, OM—outer membrane

structure to connect them (Homma and Iino 1985). The filament cap promotes the self-assembly of flagellin subunits into the filament at the distal end (Ikeda et al. 1985, 1996; Yonekura et al. 2000).

Electron cryo-microscopy (cryo-EM) image analyses of the unusually elongated rod structure, namely polyrod, which is caused by the G65V substitution in the distal rod protein FlgG, have revealed that FlgG is composed of three domains, D0, Dc, and D1 from the inside to outside of the polyrod structure (Fig. 8.3c) (Fujii et al. 2017; Saijo-Hamano et al. 2019). Intermolecular interactions between FlgG subunits in the polyrod structure are very tight, and hence the rod is straight and rigid against bending to act as a drive shaft. Because FliE is an axial component protein of the basal body and interacts with the proximal rod protein FlgB, it is thought to form the most proximal part of the rod structure that is directly connected with the MS ring and/or the transmembrane export gate complex of the type III protein export apparatus (Minamino and Macnab 2000b; Minamino et al. 2000).

The hook is a short, curved structure composed of about 120 copies of the hook protein FlgE. FlgE consist of four domains, D0, Dc, D1 and D2, arranged from the inside to the outside of the hook structure (Fig. 8.3c) (Samatey et al. 2004; Fujii et al. 2009). Although the D0, Dc and D1 domains of FlgE are highly homologous to those of FlgG, there are structural and functional differences in the Dc domain

between FlgE and FlgG (Chevance et al. 2007; Fujii et al. 2017; Saijo-Hamano et al. 2019). The Dc domain of FlgG contains a FlgG-specific YQTIRQPGAQSSEQTTLP sequence (GSS), whereas FlgE does not. Insertion of the GSS into a corresponding region of the Dc domain of FlgE makes the hook straight and rigid in a way similar to the FlgG rod structure, suggesting that the GSS is important for the rigidity of the rod structure (Hiraoka et al. 2017). Axial packing interactions between a triangular loop of the D1 domain and the D2 domain in each protofilament of the hook structure are responsible for hook supercoiling (Samatey et al. 2004; Kato et al. 2019). However, a deletion of either a triangular loop of the D1 domain or domain D2 makes the hook straight but retains its bending flexibility, allowing the filaments to form a flagellar bundle behind a cell body (Sakai et al. 2018). Because there are gaps between FlgE subunits in the entire hook structure, these gaps are likely to contribute to bending flexibility of the hook (Fujii et al. 2009; Horvath et al. 2019; Kato et al. 2019; Shibata et al. 2019).

The filament is composed of more than 30,000 flagellin subunits and is a major target for antigens. *Salmonella* contains two flagellin genes *fliC* and *fljB* on the chromosomal DNA and expresses either FliC or FljB to form the filament on the cell surface. The alternate expression frequency of these two flagellin genes is about 10^{-3}–10^{-5} per cell per generation (Stocker 1949; Gillen and Hughes 1991), and such an autonomous switching changes the motility function as well as antigen response of *Salmonella* cells (Yamaguchi et al. 2020). The flagellin subunit is composed of four domains D0, D1, D2, and D3, arranged from the inner core to the outer surface of the filament (Fig. 8.3a, c) (Namba and Vonderviszt 1997). The D0 and D1 domains are responsible for the formation of the supercoiled form of the filament structure, and a β-hairpin in the D1 domain is thought to act as a structural switch that changes between left-handed and right-handed helical forms of the filament structure (Samatey et al. 2001). Because the D3 domain, which occupies the outermost part of the filament, is recognized as H antigens by host immune systems, the primary sequence of the D3 domain differs significantly between FliC and FljB. Interestingly, there are structural differences in the D3 domain between the FliC and FljB filaments, and such differences cause a significant difference in the helical propeller function of the filament under viscous conditions (Yamaguchi et al. 2020). The D3 domain of the flagellin molecule is dispensable for filament formation but contributes to thermal stability of the entire filament structure (Muskotal et al. 2010; Furukawa et al. 2016).

The filament growth occurs at the distal end of the growing filament structure (Minamino 2014). Purified flagellin subunits can self-assemble into a long helical filament in vitro (Asakura 1970). However, flagellin subunits require a filament cap formed by FliD to polymerize into the filament structure in vivo (Homma et al. 1984). If the FliD cap is missing, flagellin monomers cannot assemble into the filament and hence are secreted into the culture media. This suggests that the FliD cap prevents newly exported flagellin molecules from leaking into the culture media to ensure that each flagellin molecule has sufficient time to be incorporated into the filament. Interestingly, the stoichiometry of the filament cap structure is different among bacteria species. The *Salmonella* FliD cap has five-fold rotational symmetry (Fig. 8.3c), the *Escherichia coli* and *Pseudomonas aeruginosa* FliD caps have six-fold rotational

symmetry, and the *Serratia marcescens* and *Bdellovibrio bacteriovorus* FliD caps form a four-fold rotational symmetry structure (Yonekura et al. 2000; Maki-Yonekura et al. 2003; Postel et al. 2016; Ko et al. 2017; Song et al. 2017; Cho et al. 2019). These differences might be related to the structural diversity of the flagellin molecule.

Rotor

The *Salmonella* basal body has the C ring (FliG, FlM, FliN), the MS ring (FliF), the P ring (FlgI) and the L ring (FlgH), which are located in the cytoplasm, the cytoplasmic membrane, the peptidoglycan (PG) layer and the outer membrane, respectively (Fig. 8.4a). The MS ring and the C ring together form a rotor ring complex of the flagellar motor. The most proximal part of the rod is thought to be directly connected with the MS ring. The L and P rings together function as a bushing for the

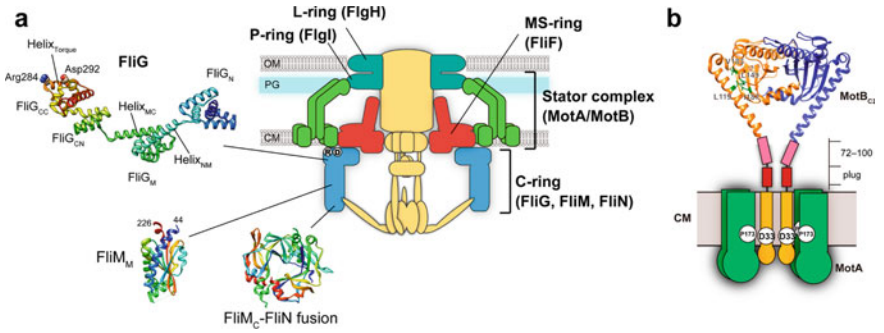


Fig. 8.4 Flagellar motor complex structure. **a** Cartoon of basal body rings. The basal body contains four ring structures named the C ring (FliG, FliM, FliN), the MS ring (FliF), the P ring (FlgI) and the L ring (FlgH). The MS ring and the C ring form the rotor ring complex of the flagellar motor. Crystal structures of *Aquifex aeolicus* FliG (PDB ID, 3HJL), *Thermotoga maritima* FliM middle domain ($FliM_M$) (PDB ID, 2HP7) and the *Salmonella* $FliM_C$-FliN fusion protein (PDB ID, 4YXB). FliG consists of three compactly folded domains, $FliG_N$, $FliG_M$ and $FliG_C$ that is able to be divided into two $FliG_{CN}$ and $FliG_{CC}$ subdomains and two linker helices, $Helix_{NM}$ connecting $FliG_N$ and $FliG_M$ and $Helix_{MC}$ connecting $FliG_M$ and $FliG_{CN}$. Two highly conserved charged residues, Arg284 (R) and Asp292 (D) in the torque helix of $FliG_{CC}$ are responsible for the interaction with the stator protein MotA. **b** Structure of the stator unit of the flagellar motor. The stator unit consists of four MotA and two MotB proteins and acts as a transmembrane proton channel to couple the proton flow to torque generation. Highly conserved Pro173 of MotA (P173) and Asp33 of MotB (D33) are directly involved in the proton translocation. The periplasmic domain of MotB ($MotB_C$) forms a homodimer and binds to the PG layer, allowing the MotA/MotB complex to act as a stator unit in the motor. A crystal structure of the $MotB_C$ dimer (PDB ID, 2ZVY) is shown. A linker region connecting MotB-TM and $MotB_C$ is dispensable for flagellar motor rotation but coordinates the proton channel activity of the MotA/MotB complex to stator assembly into the motor. The plug segment suppresses premature proton flow through the proton channel until the MotA/MotB complex become an active stator unit in the motor. CM—cytoplasmic membrane, PG—peptidoglycan layer, OM—outer membrane

distal rod structure, and the inner surface of the LP ring complex is likely to be very smooth for high-speed rotation of the rod as a drive shaft (Minamino et al. 2008).

FliF molecules with two transmembrane (TM) helices assemble into the MS ring in the cytoplasmic membrane. Over-expression of FliF alone from a pET-based plasmid vector results in the self-assembly of FliF subunits into the MS ring (Ueno et al. 1992). However, when FliF is expressed from the chromosomal DNA, FliF requires FliG for efficient formation of the MS ring in the cytoplasmic membrane (Morimoto et al. 2014). The N-terminal domain of FliG ($FliG_N$) binds to the C-terminal cytoplasmic domain of FliF (Kihara et al. 2000; Levenson et al. 2012). Intermolecular interactions between $FliG_N$ domains and between the middle ($FliG_M$) and C-terminal ($FliG_C$) domains of FliG are responsible for FliG ring formation on the cytoplasmic surface of the MS ring (Baker et al. 2016; Kim et al. 2017). The FliM/FliN complexes with a 1 FliM: 3 FliN stoichiometry bind to the FliG ring through an interaction between $FliG_M$ and FliM to form the C ring wall with 34-fold rotational symmetry (Paul et al. 2011; Vartanian et al. 2012; Lam et al. 2013). Because the torque helix in the C-terminal portion of $FliG_C$ ($FliG_{CC}$) is directly involved in the interaction with the stator protein MotA to generate the rotational force, $FliG_{CC}$ is believed to be located at the top of the C ring wall (Fig. 8.4a) (Zhou et al. 1998a). It has been reported that there is a symmetry mismatch between the MS and C rings (Suzuki et al. 2004), but recent high-resolution cryo-EM image analysis of the *Salmonella* MS ring have shown that the MS ring also has about 34-fold rotational symmetry (Johnson et al. 2020).

Many *Salmonella* mutants with altered motility have been isolated. Among motility mutants, mutations in FliG, FliM, and FliN confer a CCW switch bias phenotype or a CW switch bias phenotype even in the presence of the chemosensory signal transduction pathway. This suggests that these three C ring proteins are involved in directional switching of the flagellar motor (Yamaguchi et al. 1986). Genetic and biochemical analyses have revealed that the chemotaxis signaling protein CheY-P directly binds to the intrinsically disordered N-terminal region of FliM and the C-terminal domain of FliN in the C ring and promotes a highly cooperative conformational change in the FliG ring structure, allowing the motor to rotate in CW direction (Dyer et al. 2009; Sarkar et al. 2010). A conserved MFXF motif in a flexible hinge connecting the N-terminal portion of $FliG_C$ ($FliG_{CN}$) and $FliG_{CC}$ is thought to induce a 180° rotation of the torque helix of $FliG_{CC}$ relative to $FliG_{CN}$ when the motor switches the direction of rotation from CCW to CW (Lam et al. 2012; Miyanoiri et al. 2017). An in-frame deletion of three residues (Pro169–Ala170–Ala171) near the N-terminal portion of the linker helix connecting $FliG_M$ and $FliG_{CN}$ ($Helix_{MC}$) causes an extremely strong CW switch bias phenotype. The PAA deletion induces not only a distinct orientation of $Helix_{MC}$ relative to $FliG_M$ but also the rotation of $FliG_{CC}$ relative to $FliG_{CN}$, suggesting that $Helix_{MC}$ is postulated to act as a gear to coordinately switch the FliG ring structure between CCW and CW states (Minamino et al. 2011a; Kinoshita et al. 2018a, b). In addition to CheY-P, cyclic di-GMP (c-di-GMP), which is a second messenger molecule to induce biofilm formation, binds to YcgR to regulate chemotaxis through interactions of the c-di-GMP bound form of YcgR

with FliG and FliM in the C ring as well as MotA (Boehm et al. 2010; Fang and Gomelsky 2010; Paul et al. 2010).

The $FliM_1/FliN_3$ complex labeled with a fluorescent protein is visualized to exchange rapidly between the basal body and its cytoplasmic pool by high-resolution single molecule imaging techniques whereas neither FliF nor FliG does not, indicating that the C ring wall has a highly dynamic nature whereas both MS and FliG rings do not (Delalez et al. 2010; Fukuoka et al. 2010). The copy number of $FliM_1/FliN_3$ complex is estimated to be about 1.3 times higher in the CCW motor than in the CW motor (Delalez et al. 2014). Because the rapid turnover of the C ring wall depends on CheY-P, the directional switching of the flagellar motor seems to induce the dissociation of weakly bound $FliM_1/FliN_3$ complexes from the C ring wall (Lele et al. 2012; Branch et al. 2014; Delalez et al. 2014). The number of the $FliM_1/FliN_3$ complexes in the C ring increases with a decrease in the cytoplasmic concentration of CheY-P, suggesting that the motor adapts to changes in the steady-state level of CheY-P by tuning the copy number of the $FliM_1/FliN_3$ complex associated with the C ring wall (Lele et al. 2012; Branch et al. 2014; Delalez et al. 2014). However, because both purified CCW and CW motors have the C ring with 34-fold rotational symmetry (Sakai et al. 2019), it remains unknown how the FliG ring can accommodate additional $FliM_1/FliN_3$ complexes to the C ring.

Stator

MotA and MotB form a transmembrane proton channel that acts as a stator unit of the flagellar motor to couple the proton flow through the proton channel to torque generation (Fig. 8.4a) (Larsen et al. 1974; Manson et al. 1977; Ravid and Eisenbach 1984). MotA and MotB form a heterohexamer with a 4 MotA: 2 MotB stoichiometry and has two distinct proton channels (Fig. 8.4b) (Braun et al. 2004; Kojima and Blair 2004). MotA assembles into a homo-tetramer (Braun et al. 2004; Kim et al. 2008; Takekawa et al. 2016), and the MotB homo-dimer binds to the central pore of the MotA tetramer (Braun and Blair 2001). MotA has four TM helices (TM1–TM4), two short periplasmic loops between TM1 and TM2 and between TM3 and TM4, a relatively large cytoplasmic loop between TM2 and TM3 and a C-terminal cytoplasmic tail (Zhou et al. 1995). MotB has an N-terminal cytoplasmic tail, followed by a single TM helix and finally a large C-terminal periplasmic domain containing a peptidoglycan-binding (PGB) motif ($MotB_C$) (Kojima et al. 2009). The MotB-TM helix forms a proton channel with the TM3 and TM4 helices of MotA (Braun and Blair 2001; Braun et al. 2004). Highly conserved Asp33 of MotB and Pro173 of MotA, which are located in the proton channel, are involved in the proton flow through the channel (Zhou et al. 1998b; Braun et al. 1999; Kojima and Blair 2001; Che et al. 2008). The cytoplasmic loop of MotA contains highly conserved Arg90 and Glu98 residues that are responsible for the interaction with the torque helix of $FliG_{CC}$ (Zhou et al. 1998a). $MotB_C$ forms a dimer, and its dimerization is critical for efficient targeting and stable anchoring of the stator units to the PG layer. As a

result, the $MotA_4/MotB_2$ complex can become an active stator unit around a rotor (Kojima et al. 2009, 2018). Interestingly, site-specific disulfide bridge experiments have shown that $MotB_C$ is in relatively close proximity to the P ring of the basal body (Fig. 8.4a) (Hizukuri et al. 2010).

A flexible linker region of MotB (residues 51–100) connecting MotB-TM and the $MotB_C$ domain are dispensable for flagellar motor rotation (Fig. 8.4b) (Muramoto and Macnab 1998). Because an in-frame deletion of residues 53–66 of MotB considerably reduces cytoplasmic pH due to undesirable proton leakage through unassembled $MotA_4/MotB_2$ proton channel complex, residues 53–66 of MotB are postulated to form a plug segment to suppress the proton leakage through the proton channel until the $MotA_4/MotB_2$ complex is incorporated into a motor (Hosking et al. 2006; Morimoto et al. 2010a). Although the $MotA_4/MotB_2$ complex with a deletion of residues 51–100 of MotB is still functional, such a large deletion does not induce a considerable proton leakage through unassembled $MotA_4/MotB_2$ proton channel complexes (Muramoto and Macnab 1998). The MotB (L119P) substitution in the PGB domain of $MotB_C$ not only causes massive proton leakage though unassembled $MotA_4/MotB\ (\Delta 51–100)_2$ complexes (Kojima et al. 2009; Morimoto et al. 2010a) but also increase the binding affinity of the PGB domain for the PG layer (Kojima et al. 2018). The MotB (L119P) substitution induces an order-to-disorder transition of the N-terminal portion of the PGB domain (Kojima et al. 2018). Since a 5 nm extension of the PGB domain of MotB from the MotB-TM helix is required for the binding of the PGB domain to the PG layer (Kojima et al. 2008), such an order-to-disordered transition of the N-terminal portion of the PGB domain should occur when the $MotA_4/MotB_2$ complex encounters a rotor to become an active stator unit in the motor (Kojima et al. 2018). In agreement with this, the 5 nm extension process of the PGB domain of MotS derived from *Bacillus subtilis*, which is a MotB homologue, has been directly visualized using high-speed atomic force microscopy (Terahara et al. 2017a). Therefore, these observations suggest that the flexible linker region of MotB coordinates not only proper proton channel formation of the $MotA_4/MotB_2$ complex but also efficient and proper anchoring of the PGB domain to the PG layer.

Each stator complex binds to and dissociates from a rotor ring complex during torque generation by the flagellar motor, suggesting that the stator complex has a highly dynamic nature (Leake et al. 2006). The number of functionally active stators in the motor varies in response to changes in the environment, such as external load and external proton concentration (Fukuoka et al. 2009; Lele et al. 2013; Tipping et al. 2013; Che et al. 2014; Terahara et al. 2017b; Suzuki et al. 2019). For example, the maximum number of active stator units incorporated into the flagellar motor is about ten when the motor operates at high loads (Reid et al. 2006; Pourjaberi et al. 2017). In contrast, the stator number decreases from ten to a few with a decrease in external load (Yuan and Berg 2008; Nakamura et al. 2020). Thus, the flagellar motor autonomously controls the number of active stator units around the rotor in response to changes in external loads. It has been shown that an in-frame deletion of residues 72–100 of MotB results in much steeper decrease in the stator number with a decrease in the external load as compared to the wild-type motor (Castillo et al. 2013). This suggests that the linker region of MotB modulates the binding affinity of

the PGB domain for the PG layer in a load-dependent manner. Interestingly, certain point mutations in the cytoplasmic loop of MotA also change the load-sensitivity of the $MotA_4/MotB_2$ complex (Pourjaberi et al. 2017). Because the cytoplasmic loop of MotA directly binds to FliG to generate torque (Zhou et al. 1998a; Morimoto et al. 2010b; Morimoto et al. 2013), this cytoplasmic loop seems to act as a load sensor to coordinate the number of active stators incorporated into the motor in response to changes in external load. Recently, it has been shown that the localization efficiency of the $MotA_4/MotB_2$ complex is also increased by lowering extracellular pH, that is, increasing proton concentration, and that such an external pH-dependent localization pattern of the $MotA_4/MotB_2$ complex is not affected by the *motB(D33N)* mutation, which inhibits the proton flow through the $MotA_4/MotB_2$ proton channel complex (Suzuki et al. 2019). Since it has been reported that the PGB domain of MotS induces a disordered-to-order transition of the PGB domain in a Na^+-dependent manner to allow the $MotP_4/MotS_2$ complex to become an active stator unit in the *B. subtilis* flagellar motor (Terahara et al. 2017a), it is possible that the folding efficiency of the PGB domain of MotB depends on external pH values.

Type III Protein Export Apparatus

The assembly of the axial structure begins with the rod, followed by the hook, the junction, the filament cap in this order and finally the filament (Macnab 2003). For the self-assembly of each building block at the distal end of the growing structure, a type III protein export apparatus unfolds the building blocks synthesized in the cytoplasm and then transports them into the central channel of the growing structure (Minamino 2014; Minamino et al. 2020a). PMF across the cytoplasmic membrane and ATP are utilized as the energy sources to drive flagellar protein export by the type III protein export apparatus (Minamino and Namba 2008; Paul et al. 2008). The type III protein export apparatus is composed of a PMF-driven export gate complex consisting of five membrane proteins FlhA, FlhB, FliP, FliQ and FliR, and a cytoplasmic ATPase ring complex made of three cytoplasmic proteins, FliH, FliI and FliJ (Fig. 8.5) (Minamino 2014). These component proteins show sequence and functional similarities to those of the injectisome of pathogenic bacteria such as *Salmonella* spp, *Shigella* spp and *Yersinia* spp, which is involved in direct injection of effector proteins into eukaryotic host cells for bacterial infection (Galán et al. 2014). Furthermore, the overall structure of the cytoplasmic ATPase ring complex is similar to F-type and V-type rotary ATPase families, suggesting that the type III protein export apparatus of the flagellum and the injectisome share a common evolutionary origin with these rotary ATPase families (Imada et al. 2007, 2016; Ibuki et al. 2011). In addition, four cytoplasmic proteins, FlgN, FliA, FliS and FliT, function as flagellum-specific export chaperones to support the export of their cognate building blocks by the type III protein export apparatus (Table 8.1).

FliP, FliQ, FliR form a helical assembly with a 5 FliP: 4 FliQ: 1 FliR stoichiometry inside the central pore of the MS ring (Fig. 8.5) (Fabiani et al. 2017; Fukumura

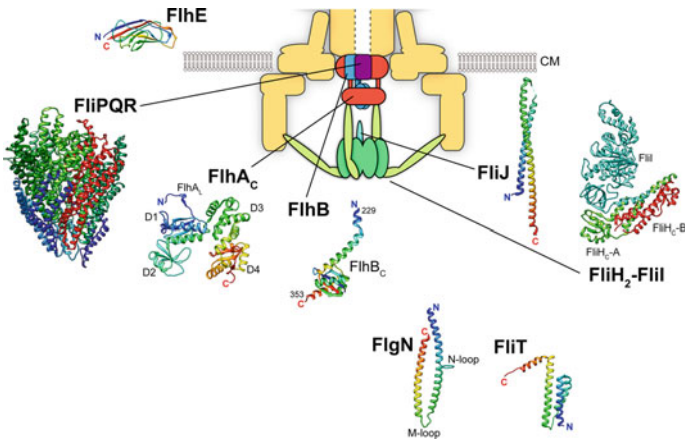


Fig. 8.5 Structures of the flagellar type III export apparatus component proteins. The type III protein export apparatus is composed of five transmembrane proteins, FlhA, FlhB, FliP, FliQ and FliR and three cytoplasmic protein FliH, FliI and FliJ. In addition, flagellar export chaperones such as FlgN and FliT interact with the type III protein export apparatus proteins to promote the export of their cognate building blocks. Crystal structures of the C-terminal cytoplasmic domain of *Salmonella* FlhA (FlhA$_C$) (PDB ID, 3A5I), the C-terminal cytoplasmic domain of *Salmonella* FlhB (FlhB$_C$) (PDB ID, 3B0Z), the *Salmonella* FliP/FliQ/FliR complex with a 5 FliP (green): 4 FliQ (blue): 1 FliR (red) stoichiometry (PDB ID, 6R69), *Salmonella* FlhE (PDB ID, 4QXL), *Salmonella* FliH$_2$FliI complex with a 2 FliH: 1 FliI stoichiometry (PDB ID, 5B0O), *Salmonella* FliJ (PDB ID, 3AJW) and the *Salmonella* FlgN (PDB ID, 5B3D) and FliT (PDB ID, 3A7M) export chaperones are shown. FlhA$_C$ consists of four compactly folded domains, D1, D2, D3 and D4 and a flexible linker connecting FlhA$_C$ and the N-terminal transmembrane region of FlhA. Hinge loops of the FlgN and FliT chaperones are responsible for helical rearrangements of these chaperone structure to modulate the binding affinity for their binding partners. CM—cytoplasmic membrane

et al. 2017; Kuhlen et al. 2018; Butan et al. 2019). The FliP$_5$/FliQ$_4$/FliR$_1$ complex is thought to be connected with the most proximal part of the rod to form a continuous protein channel inside the growing axial structure (Kuhlen et al. 2018). FlhA forms a homo-nonamer through intermolecular interactions between its C-terminal cytoplasmic domains (FlhA$_C$) (Abrusci et al. 2013; Kawamoto et al. 2013; Morimoto et al. 2014; Terahara et al. 2018) and acts as an energy transducer along with the cytoplasmic ATPase complex (Minamino et al. 2011b, 2016b; Erhardt et al. 2017). FlhA associates not only with the FliP$_5$/FliQ$_4$/FliR$_1$ helical structure but also with the MS ring (Kihara et al. 2001; Fukumura et al. 2017). A single copy of FlhB associates with FliP and FliR (Kuhlen et al. 2019). The assembly of the export gate complex begins with the formation of the FliP$_5$/FliR$_1$ complex with the help of the transmembrane protein FliO, followed by the assembly of FliQ and FlhB presumably in this order and finally the assembly of FlhA (Morimoto et al. 2014; Fabiani et al. 2017; Fukumura et al. 2017; Kuhlen et al. 2018). FlhA$_C$ and the C-terminal cytoplasmic domain of FlhB (FlhB$_C$) project into the central cavity of the C ring (Fig. 8.5) and form a docking platform for the cytoplasmic ATPase complex, flagellar export chaperones and building blocks (Bange et al. 2010; Minamino et al. 2010, 2012). These two

$FlhA_C$ and $FlhB_C$ domains mediate well-organized protein targeting and transport for efficient assembly of the axial structure with the help of the cytoplasmic ATPase complex and the molecular ruler protein FliK (Table 8.1) (Minamino and Macnab 2000a; Fraser et al. 2003; Hirano et al. 2009; Kinoshita et al. 2013; Minamino et al. 2016a, 2020b; Inoue et al. 2018, 2019).

FliI is a member of the Walker-type ATPase family (Fan and Macnab 1996) and forms a homo-hexamer to hydrolyze ATP at an interface between FliI subunits in a way similar to the $\alpha_3\beta_3$ hetero-hexameric ring complex of F_1-ATPase (Claret et al. 2003; Imada et al. 2007; Minamino et al. 2006). FliJ binds to the center of the FliI hexamer in a way similar to the γ subunit of the F_1-ATPase (Ibuki et al. 2011). The FliH dimer, which structurally looks similar to the peripheral stalk of the F_1-ATPase (Imada et al. 2016), binds to the N-terminal domain of each FliI subunit (Okabe et al. 2009) and anchors the $FliI_6FliJ$ ring complex to the flagellar base through interactions of the extreme N-terminal region of FliH with FliN and FlhA (Bai et al. 2014; Gonzalez-Pedrajo et al. 2006; Minamino et al. 2009; Hara et al. 2012). ATP hydrolysis by the FliI ATPase activates the PFM-driven export gate complex through an interaction between the FliJ stalk and FlhA, allowing the export gate complex to transport building blocks from the cytoplasm to the distal end of the growing structure in a PMF-dependent manner (Minamino et al. 2011b; Ibuki et al. 2013). The FliI ATPase shows rapid exchanges between the basal body and the cytoplasmic pool, suggesting that the cytoplasmic ATPase complex seems to be a dynamic structure (Bai et al. 2014).

When the cytoplasmic ATPase complex is missing, the transmembrane export gate complex utilizes sodium motive force across the cell membrane in addition to PFM (Minamino et al. 2016b). FlhA acts as the transmembrane ion channel of the export gate complex to conduct not only protons but also sodium ions (Minamino et al. 2016b). The transmembrane export gate complex functions as a protein/cation antiporter that couples the inward-directed cation flow through the FlhA channel to the outward-directed protein transport (Minamino et al. 2011b). The *flhE* gene forms an operon together with the *flhB* and *flhA* genes, and its gene product is secreted via the Sec translocon into the periplasm (Minamino et al. 1994). FlhE is postulated to act as a plug to prevent the FlhA ion channel from translocating large numbers of protons along an electrochemical difference of protons across the cytoplasmic membrane (Lee et al. 2015), but it remains unknown how it works. High-resolution local pH measurements around the type III protein export apparatus have shown that a loss of the ATPase ring complex increases the local pH inside the C ring, indicating ATP hydrolysis by the ATPase ring complex facilitates the proton channel activity of the FlhA channel coupled with flagellar protein export (Morimoto et al. 2016).

The FlgN, FliS and FliT chaperones not only suppresses premature aggregation and/or proteolysis of their cognate building blocks in the cytoplasm (Auvray et al. 2001; Bennett et al. 2001; Aldridge et al. 2003) but also facilitate their docking to the type III protein export apparatus (Thomas et al. 2004; Evans et al. 2006; Kinoshita et al. 2013). The FlgN, FliS and FliT chaperones adopt an α-helical structure (Fig. 8.5) (Evdokimov et al. 2003; Imada et al. 2010; Kinoshita et al. 2016), and the binding of their cognate building block to these chaperones induces helical rearrangements

of the chaperones through flexible hinge loops, allowing the building blocks to be efficiently transferred to the type III protein export apparatus, followed by efficient unfolding and the subsequent translocation of the building blocks by the PMF-driven export gate complex (Imada et al. 2010; Kinoshita et al. 2013, 2016; Furukawa et al. 2016).

Conclusions

Detailed domain structures of bacterial flagellar component proteins have been revealed at an atomic resolution by X-ray crystallography. In addition, cryo-EM image analyses of the bacterial flagellum have provided deep insights into the mechanical functions of the rod, the hook and the filament. Cryo-electron tomographic techniques have revealed the in situ structures of the transmembrane part of the basal body structure. However, while static structural knowledge is increasing, there are still many mysterious points that do not connect to dynamic functions of the bacterial flagellum. Biophysical and molecular genetic approaches combined with structural studies are required to more accurately investigate the dynamic mechanism of the flagellar motor complex.

Acknowledgements We acknowledge Profs. Keiichi Namba and Takuo Yasunaga for continuous support and encouragement. Our research is supported in part by the Japan Society for Promotion and Science.

References

Abrusci P, Vergara-Irigaray M, Johnson S, Beeby MD, Hendrixson DR, Roversi P, Friede ME, Deane JE, Jensen GJ, Tang CM, Lea SM (2013) Architecture of the major component of the type III secretion system export apparatus. Nat Struct Mol Biol 20(1):99–104. https://doi.org/10.1038/nsmb.2452

Aldridge P, Karlinsey J, Hughes KT (2003) The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. Mol Microbiol 49(5):1333–1345. https://doi.org/10.1046/j.1365-2958.2003.03637.x

Asakura S (1970) Polymerization of flagellin and polymorphism of flagella. Adv Biophys 1:99–155

Auvray F, Thomas J, Fraser GM, Hughes C (2001) Flagellin polymerisation control by a cytosolic export chaperone. J Mol Biol 308(2):221–229. https://doi.org/10.1006/jmbi.2001.4597

Bai F, Morimoto YV, Yoshimura SDJ, Hara N, Kami-Ike N, Namba K, Minamino T (2014) Assembly dynamics and the roles of FliI ATPase of the bacterial flagellar export apparatus. Sci Rep 4:6528. https://doi.org/10.1038/srep06528

Baker MA, Hynson RM, Ganuelas LA, Mohammadi NS, Liew CW, Rey AA, Duff AP, Whitten AE, Jeffries CM, Delalez NJ, Morimoto YV, Stock D, Armitage JP, Turberfield AJ, Namba K, Berry RM, Lee LK (2016) Domain-swap polymerization drives the self-assembly of the bacterial flagellar motor. Nat Struct Mol Biol 23(3):197–203. https://doi.org/10.1038/nsmb.3172

Bange G, Kummerer N, Engel C, Bozkurt G, Wild K, Sinning I (2010) FlhA provides the adaptor for coordinated delivery of late flagella building blocks to the type III secretion system. Proc Natl Acad Sci USA 107(25):11295–11300. https://doi.org/10.1073/pnas.1001383107

Bennett JC, Thomas J, Fraser GM, Hughes C (2001) Substrate complexes and domain organization of the *Salmonella* flagellar export chaperones FlgN and FliT. Mol Microbiol 39(3):781–791. https://doi.org/10.1046/j.1365-2958.2001.02268.x

Berg HC (2003) The rotary motor of bacterial flagella. Annu Rev Biochem 72:19–54. https://doi.org/10.1146/annurev.biochem.72.121801.161737

Bi S, Sourjik V (2018) Stimulus sensing and signal processing in bacterial chemotaxis. Curr Opin Microbiol 45:22–29. https://doi.org/10.1016/j.mib.2018.02.002

Boehm A, Kaiser M, Li H, Spangler C, Kasper CA, Ackermann M, Kaever V, Sourjik V, Roth V, Jenal U (2010) Second messenger-mediated adjustment of bacterial swimming velocity. Cell 141(1):107–116. https://doi.org/10.1016/j.cell.2010.01.018

Branch RW, Sayegh MN, Shen C, Nathan VSJ, Berg HC (2014) Adaptive remodelling by FliN in the bacterial rotary motor. J Mol Biol 426(19):3314–3324. https://doi.org/10.1016/j.jmb.2014.07.009

Braun TF, Blair DF (2001) Targeted disulfide cross-linking of the MotB protein of *Escherichia coli*: evidence for two H^+ channels in the stator complex. Biochemistry 40(43):13051–13059. https://doi.org/10.1021/bi011264g

Braun TF, Poulson S, Gully JB, Empey JC, Van Way S, Putnam A, Blair DF (1999) Function of proline residues of MotA in torque generation by the flagellar motor of *Escherichia coli*. J Bacteriol 181(11):3542–3551

Braun TF, Al-Mawsawi LQ, Kojima S, Blair DF (2004) Arrangement of core membrane segments in the MotA/MotB proton-channel complex of *Escherichia coli*. Biochemistry 43(1):35–45. https://doi.org/10.1021/bi035406d

Butan C, Lara-Tejero M, Li W, Liu J, Galan JE (2019) High-resolution view of the type III secretion export apparatus in situ reveals membrane remodeling and a secretion pathway. Proc Natl Acad Sci USA 116(49):24786–24795. https://doi.org/10.1073/pnas.1916331116

Calladine CR (1975) Construction of bacterial flagella. Nature 255(5504):121–124. https://doi.org/10.1038/255121a0

Calladine CR (1976) Design requirements for the construction of bacterial flagella. J Theor Biol 57(2):469–489. https://doi.org/10.1016/0022-5193(76)90016-3

Castillo DJ, Nakamura S, Morimoto YV, Che YS, Kami-Ike N, Kudo S, Minamino T, Namba K (2013) The C-terminal periplasmic domain of MotB is responsible for load-dependent control of the number of stators of the bacterial flagellar motor. Biophysics 9:173–181. https://doi.org/10.2142/biophysics.9.173

Che YS, Nakamura S, Kojima S, Kami-ike N, Namba K, Minamino T (2008) Suppressor analysis of the MotB(D33E) mutation to probe bacterial flagellar motor dynamics coupled with proton translocation. J Bacteriol 190(20):6660–6667. https://doi.org/10.1128/JB.00503-08

Che YS, Nakamura S, Morimoto YV, Kami-Ike N, Namba K, Minamino T (2014) Load-sensitive coupling of proton translocation and torque generation in the bacterial flagellar motor. Mol Microbiol 91(1):175–184. https://doi.org/10.1111/mmi.12453

Chevance FF, Takahashi N, Karlinsey JE, Gnerer J, Hirano T, Samudrala R, Aizawa S, Hughes KT (2007) The mechanism of outer membrane penetration by the eubacterial flagellum and implications for spirochete evolution. Genes Dev 21(18):2326–2335. https://doi.org/10.1101/gad.1571607

Cho SY, Song WS, Yoon SI (2019) Crystal structure of the flagellar cap protein FliD from *Bdellovibrio bacteriovorus*. Biochem Biophys Res Commun 519(3):652–658. https://doi.org/10.1016/j.bbrc.2019.09.024

Claret L, Calder SR, Higgins M, Hughes C (2003) Oligomerization and activation of the FliI ATPase central to bacterial flagellum assembly. Mol Microbiol 48(5):1349–1355. https://doi.org/10.1046/j.1365-2958.2003.03506.x

Delalez NJ, Wadhams GH, Rosser G, Xue Q, Brown MT, Dobbie IM, Berry RM, Leake MC, Armitage JP (2010) Signal-dependent turnover of the bacterial flagellar switch protein FliM. Proc Natl Acad Sci USA 107(25):11347–11351. https://doi.org/10.1073/pnas.1000284107

Delalez NJ, Berry RM, Armitage JP (2014) Stoichiometry and turnover of the bacterial flagellar switch protein FliN. mBio 5(4):e01216–01214. https://doi.org/10.1128/mBio.01216-14

Dyer CM, Vartanian AS, Zhou H, Dahlquist FW (2009) A molecular mechanism of bacterial flagellar motor switching. J Mol Biol 388(1):71–84. https://doi.org/10.1016/j.jmb.2009.02.004

Erhardt M, Wheatley P, Kim EA, Hirano T, Zhang Y, Sarkar MK, Hughes KT, Blair DF (2017) Mechanism of type-III protein secretion: regulation of FlhA conformation by a functionally critical charged-residue cluster. Mol Microbiol 104(2):234–249. https://doi.org/10.1111/mmi.13623

Evans LD, Stafford GP, Ahmed S, Fraser GM, Hughes C (2006) An escort mechanism for cycling of export chaperones during flagellum assembly. Proc Natl Acad Sci USA 103(46):17474–17479. https://doi.org/10.1073/pnas.0605197103

Evdokimov AG, Phan J, Tropea JE, Routzahn KM, Peters HK, Pokross M, Waugh DS (2003) Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion. Nat Struct Biol 10(10):789–793. https://doi.org/10.1038/nsb982

Fabiani FD, Renault TT, Peters B, Dietsche T, Galvez EJC, Guse A, Freier K, Charpentier E, Strowig T, Franz-Wachtel M, Macek B, Wagner S, Hensel M, Erhardt M (2017) A flagellum-specific chaperone facilitates assembly of the core type III export apparatus of the bacterial flagellum. PLoS Biol 15(8):e2002267. https://doi.org/10.1371/journal.pbio.2002267

Fan F, Macnab RM (1996) Enzymatic characterization of FliI—an ATPase involved in flagellar assembly in *Salmonella* typhimurium. J Biol Chem 271(50):31981–31988

Fang X, Gomelsky M (2010) A post-translational, c-di-GMP-dependent mechanism regulating flagellar motility. Mol Microbiol 76(5):1295–1305. https://doi.org/10.1111/j.1365-2958.2010.07179.x

Fraser GM, Hirano T, Ferris HU, Devgan LL, Kihara M, Macnab RM (2003) Substrate specificity of type III flagellar protein export in *Salmonella* is controlled by subdomain interactions in FlhB. Mol Microbiol 48(4):1043–1057. https://doi.org/10.1046/j.1365-2958.2003.03487.x

Fujii T, Kato T, Namba K (2009) Specific arrangement of alpha-helical coiled coils in the core domain of the bacterial flagellar hook for the universal joint function. Structure 17(11):1485–1493. https://doi.org/10.1016/j.str.2009.08.017

Fujii T, Kato T, Hiraoka KD, Miyata T, Minamino T, Chevance FF, Hughes KT, Namba K (2017) Identical folds used for distinct mechanical functions of the bacterial flagellar rod and hook. Nat Commun 8:14276. https://doi.org/10.1038/ncomms14276

Fukumura T, Makino F, Dietsche T, Kinoshita M, Kato T, Wagner S, Namba K, Imada K, Minamino T (2017) Assembly and stoichiometry of the core structure of the bacterial flagellar type III export gate complex. PLoS Biol 15(8):e2002281. https://doi.org/10.1371/journal.pbio.2002281

Fukuoka H, Wada T, Kojima S, Ishijima A, Homma M (2009) Sodium-dependent dynamic assembly of membrane complexes in sodium-driven flagellar motors. Mol Microbiol 71(4):825–835. https://doi.org/10.1111/j.1365-2958.2008.06569.x

Fukuoka H, Inoue Y, Terasawa S, Takahashi H, Ishijima A (2010) Exchange of rotor components in functioning bacterial flagellar motor. Biochem Biophys Res Commun 394(1):130–135. https://doi.org/10.1016/j.bbrc.2010.02.129

Furukawa Y, Imada K, Vonderviszt F, Matsunami H, Sano K, Kutsukake K, Namba K (2002) Interactions between bacterial flagellar axial proteins in their monomeric state in solution. J Mol Biol 318(3):889–900. https://doi.org/10.1016/S0022-2836(02)00139-0

Furukawa Y, Inoue Y, Sakaguchi A, Mori Y, Fukumura T, Miyata T, Namba K, Minamino T (2016) Structural stability of flagellin subunit affects the rate of flagellin export in the absence of FliS chaperone. Mol Microbiol 102(3):405–416. https://doi.org/10.1111/mmi.13469

Galán JE, Lara-Tejero M, Marlovits TC, Wagner S (2014) Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. Annu Rev Microbiol 68:415–438. https://doi.org/10.1146/annurev-micro-092412-155725

Gillen KL, Hughes KT (1991) Negative regulatory loci coupling flagellin synthesis to flagellar assembly in *Salmonella typhimurium*. J Bacteriol 173(7):2301–2310. https://doi.org/10.1128/jb.173.7.2301-2310.1991

Gonzalez-Pedrajo B, Minamino T, Kihara M, Namba K (2006) Interactions between C ring proteins and export apparatus components: a possible mechanism for facilitating type III protein export. Mol Microbiol 60(4):984–998. https://doi.org/10.1016/j.1365-2958.2006.05149.x

Hara N, Morimoto YV, Kawamoto A, Namba K, Minamino T (2012) Interaction of the extreme N-terminal region of FliH with FlhA is required for efficient bacterial flagellar protein export. J Bacteriol 194(19):5353–5360. https://doi.org/10.1128/JB.01028-12

Hirano T, Mizuno S, Aizawa S, Hughes KT (2009) Mutations in *flk*, *flgG*, *flhA*, and *flhE* that affect the flagellar type III secretion specificity switch in *Salmonella enterica*. J Bacteriol 191(12):3938–3949. https://doi.org/10.1128/JB.01811-08

Hiraoka KD, Morimoto YV, Inoue Y, Fujii T, Miyata T, Makino F, Minamino T, Namba K (2017) Straight and rigid flagellar hook made by insertion of the FlgG specific sequence into FlgE. Sci Rep 7:46723. https://doi.org/10.1038/srep46723

Hizukuri Y, Kojima S, Homma M (2010) Disulphide cross-linking between the stator and the bearing components in the bacterial flagellar motor. J Biochem 148(3):309–318. https://doi.org/10.1093/jb/mvq067

Homma M, Iino T (1985) Locations of hook-associated proteins in flagellar structures of *Salmonella typhimurium*. J Bacteriol 162(1):183–189

Homma M, Fujita H, Yamaguchi S, Iino T (1984) Excretion of unassembled flagellin by *Salmonella typhimurium* mutants deficient in hook-associated proteins. J Bacteriol 159(3):1056–1059

Horvath P, Kato T, Miyata T, Namba K (2019) Structure of *Salmonella* flagellar hook reveals intermolecular domain interactions for the universal joint function. Biomolecules 9(9):462. https://doi.org/10.3390/biom9090462

Hosking ER, Vogt C, Bakker EP, Manson MD (2006) The *Escherichia coli* MotAB proton channel unplugged. J Mol Biol 364(5):921–937. https://doi.org/10.1016/j.jmb.2006.09.035

Ibuki T, Imada K, Minamino T, Kato T, Miyata T, Namba K (2011) Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. Nat Struct Mol Biol 18(3):277–282. https://doi.org/10.1038/nsmb.1977

Ibuki T, Uchida Y, Hironaka Y, Namba K, Imada K, Minamino T (2013) Interaction between FliJ and FlhA, components of the bacterial flagellar type III export apparatus. J Bacteriol 195(3):466–473. https://doi.org/10.1128/JB.01711-12

Ikeda T, Asakura S, Kamiya R (1985) "Cap" on the tip of *Salmonella* flagella. J Mol Biol 184(4):735–737. https://doi.org/10.1016/0022-2836(85)90317-1

Ikeda T, Oosawa K, Hotani H (1996) Self-assembly of the filament capping protein, FliD, of bacterial flagella into an annular structure. J Mol Biol 259(4):679–686. https://doi.org/10.1006/jmbi.1996.0349

Imada K, Minamino T, Tahara A, Namba K (2007) Structural similarity between the flagellar type III ATPase FliI and F1-ATPase subunits. Proc Natl Acad Sci USA 104(2):485–490. https://doi.org/10.1073/pnas.0608090104

Imada K, Minamino T, Kinoshita M, Furukawa Y, Namba K (2010) Structural insight into the regulatory mechanisms of interactions of the flagellar type III chaperone FliT with its binding partners. Proc Natl Acad Sci USA 107(19):8812–8817. https://doi.org/10.1073/pnas.1001866107

Imada K, Minamino T, Uchida Y, Kinoshita M, Namba K (2016) Insight into the flagella type III export revealed by the complex structure of the type III ATPase and its regulator. Proc Natl Acad Sci USA 113(13):3633–3638. https://doi.org/10.1073/pnas.1524025113

Inoue Y, Morimoto YV, Namba K, Minamino T (2018) Novel insights into the mechanism of well-ordered assembly of bacterial flagellar proteins in *Salmonella*. Sci Rep 8(1):1787. https://doi.org/10.1038/s41598-018-20209-3

Inoue Y, Ogawa Y, Kinoshita M, Terahara N, Shimada M, Kodera N, Ando T, Namba K, Kitao A, Imada K, Minamino T (2019) Structural insights into the substrate specificity switch mechanism

of the type III protein export apparatus. Structure 27(6):965-976.e6. https://doi.org/10.1016/j.str.2019.03.017

Johnson S, Fong YH, Deme J, Furlong E, Kuhlen L, Lea SM (2020) Symmetry mismatch in the MS-ring of the bacterial flagellar rotor explains the structural coordination of secretion and rotation. Nat Microbiol 5(7): 966–975. https://doi.org/10.1038/s41564-020-0703-3.

Kato T, Makino F, Miyata T, Horvath P, Namba K (2019) Structure of the native supercoiled flagellar hook as a universal joint. Nat Commun 10(1):5295. https://doi.org/10.1038/s41467-019-13252-9

Kawamoto A, Morimoto YV, Miyata T, Minamino T, Hughes KT, Kato T, Namba K (2013) Common and distinct structural features of *Salmonella* injectisome and flagellar basal body. Sci Rep 3:3369. https://doi.org/10.1038/srep03369

Kihara M, Miller GU, Macnab RM (2000) Deletion analysis of the flagellar switch protein FliG of *Salmonella*. J Bacteriol 182(11):3022–3028. https://doi.org/10.1128/jb.182.11.3022-3028.2000

Kihara M, Minamino T, Yamaguchi S, Macnab RM (2001) Intergenic suppression between the flagellar MS ring protein FliF of *Salmonella* and FlhA, a membrane component of its export apparatus. J Bacteriol 183(5):1655–1662. https://doi.org/10.1128/jb.183.5.1655-1662.2001

Kim EA, Price-Carter M, Carlquist WC, Blair DF (2008) Membrane segment organization in the stator complex of the flagellar motor: implications for proton flow and proton-induced conformational change. Biochemistry 47(43):11332–11339. https://doi.org/10.1021/bi801347a

Kim EA, Panushka J, Meyer T, Carlisle R, Baker S, Ide N, Lynch M, Crane BR, Blair DF (2017) Architecture of the flagellar switch complex of *Escherichia coli*: conformational plasticity of FliG and implications for adaptive remodeling. J Mol Biol 429(9):1305–1320. https://doi.org/10.1016/j.jmb.2017.02.014

Kinoshita M, Hara N, Imada K, Namba K, Minamino T (2013) Interactions of bacterial flagellar chaperone-substrate complexes with FlhA contribute to co-ordinating assembly of the flagellar filament. Mol Microbiol 90(6):1249–1261. https://doi.org/10.1111/mmi.12430

Kinoshita M, Nakanishi Y, Furukawa Y, Namba K, Imada K, Minamino T (2016) Rearrangements of alpha-helical structures of FlgN chaperone control the binding affinity for its cognate substrates during flagellar type III export. Mol Microbiol 101(4):656–670. https://doi.org/10.1111/mmi.13415

Kinoshita M, Furukawa Y, Uchiyama S, Imada K, Namba K, Minamino T (2018a) Insight into adaptive remodeling of the rotor ring complex of the bacterial flagellar motor. Biochem Biophys Res Commun 496(1):12–17. https://doi.org/10.1016/j.bbrc.2017.12.118

Kinoshita M, Namba K, Minamino T (2018b) Effect of a clockwise-locked deletion in FliG on the FliG ring structure of the bacterial flagellar motor. Genes Cells 23(3):241–247. https://doi.org/10.1111/gtc.12565

Kojima S, Blair DF (2001) Conformational change in the stator of the bacterial flagellar motor. Biochemistry 40(43):13041–13050. https://doi.org/10.1021/bi011263o

Kojima S, Blair DF (2004) Solubilization and purification of the MotA/MotB complex of *Escherichia coli*. Biochemistry 43(1):26–34. https://doi.org/10.1021/bi0354051

Kojima S, Furukawa Y, Matsunami H, Minamino T, Namba K (2008) Characterization of the periplasmic domain of MotB and implications for its role in the stator assembly of the bacterial flagellar motor. J Bacteriol 190(9):3314–3322. https://doi.org/10.1128/JB.01710-07

Kojima S, Imada K, Sakuma M, Sudo Y, Kojima C, Minamino T, Homma M, Namba K (2009) Stator assembly and activation mechanism of the flagellar motor by the periplasmic region of MotB. Mol Microbiol 73(4):710–718. https://doi.org/10.1111/j.1365-2958.2009.06802.x

Kojima S, Takao M, Almira G, Kawahara I, Sakuma M, Homma M, Kojima C, Imada K (2018) The helix rearrangement in the periplasmic domain of the flagellar stator B subunit activates peptidoglycan binding and ion influx. Structure 26(4):590–598, e595. https://doi.org/10.1016/j.str.2018.02.016

Ko W, Lim S, Lee W, Kim Y, Berg HC, Peskin CS (2017) Modeling polymorphic transformation of rotating bacterial flagella in a viscous fluid. Phys Rev E 95(6–1):063106. https://doi.org/10.1103/PhysRevE.95.063106

Kuhlen L, Abrusci P, Johnson S, Gault J, Deme J, Caesar J, Dietsche T, Mebrhatu MT, Ganief T, Macek B, Wagner S, Robinson CV, Lea SM (2018) Structure of the core of the type III secretion system export apparatus. Nat Struct Mol Biol 25(7):583–590. https://doi.org/10.1038/s41594-018-0086-9

Kuhlen L, Johnson S, Zeitler A, Bäurle S, Deme JC, Debo R, Fisher J, Wagner S, Lea SM (2019) The flagellar substrate specificity switch protein FlhB assembles onto the extra-membrane export gate to regulate type three secretion. bioRxiv:686782. https://doi.org/10.1101/686782

Lam KH, Ip WS, Lam YW, Chan SO, Ling TKW, Au SWN (2012) Multiple conformations of the FliG C-terminal domain provide insight into flagellar motor switching. Structure 20(2):315–325. https://doi.org/10.1016/j.str.2011.11.020

Lam KH, Lam WW, Wong JY, Chan LC, Kotaka M, Ling TK, Jin DY, Ottemann KM, Au SW (2013) Structural basis of FliG-FliM interaction in *Helicobacter pylori*. Mol Microbiol 88(4):798–812. https://doi.org/10.1111/mmi.12222

Larsen SH, Adler J, Gargus JJ, Hogg RW (1974) Chemomechanical coupling without ATP: the source of energy for motility and chemotaxis in bacteria. Proc Natl Acad Sci USA 71(4):1239–1243. https://doi.org/10.1073/pnas.71.4.1239

Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM, Armitage JP (2006) Stoichiometry and turnover in single, functioning membrane protein complexes. Nature 443(7109):355–358. https://doi.org/10.1038/nature05135

Lee J, Monzingo AF, Keatinge-Clay AT, Harshey RM (2015) Structure of *Salmonella* FlhE, conserved member of a flagellar type III secretion operon. J Mol Biol 427(6):1254–1262. https://doi.org/10.1016/j.jmb.2014.11.022

Lele PP, Branch RW, Nathan VS, Berg HC (2012) Mechanism for adaptive remodeling of the bacterial flagellar switch. Proc Natl Acad Sci USA 109(49):20018–20022. https://doi.org/10.1073/pnas.1212327109

Lele PP, Hosu BG, Berg HC (2013) Dynamics of mechanosensing in the bacterial flagellar motor. Proc Natl Acad Sci USA 110(29):11839–11844. https://doi.org/10.1073/pnas.1305885110

Levenson R, Zhou H, Dahlquist FW (2012) Structural insights into the interaction between the bacterial flagellar motor proteins FliF and FliG. Biochemistry 51(25):5052–5060. https://doi.org/10.1021/bi3004582

Macnab RM (2003) How bacteria assemble flagella. Annu Rev Microbiol 57:77–100. https://doi.org/10.1146/annurev.micro.57.030502.090832

Maki-Yonekura S, Yonekura K, Namba K (2003) Domain movements of HAP2 in the cap-filament complex formation and growth process of the bacterial flagellum. Proc Natl Acad Sci USA 100(26):15528–15533. https://doi.org/10.1073/pnas.2534343100

Maki-Yonekura S, Yonekura K, Namba K (2010) Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. Nat Struct Mol Biol 17(4):417–422. https://doi.org/10.1038/nsmb.1774

Manson MD, Tedesco P, Berg HC, Harold FM, Van der Drift C (1977) A protonmotive force drives bacterial flagella. Proc Natl Acad Sci USA 74(7):3060–3064. https://doi.org/10.1073/pnas.74.7.3060

Minamino T (2014) Protein export through the bacterial flagellar type III export pathway. Biochim Biophys Acta 1843(8):1642–1648. https://doi.org/10.1016/j.bbamcr.2013.09.005

Minamino T (2018) Hierarchical protein export mechanism of the bacterial flagellar type III protein export apparatus. FEMS Microbiol Lett 365(12):fny117. https://doi.org/10.1093/femsle/fny117

Minamino T, Namba K (2008) Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export. Nature 451(7177):485–488. https://doi.org/10.1038/nature06449

Minamino T, Macnab RM (2000a) Domain structure of *Salmonella* FlhB, a flagellar export component responsible for substrate specificity switching. J Bacteriol 182(17):4906–4914. https://doi.org/10.1128/jb.182.17.4906-4914.2000

Minamino T, Macnab RM (2000b) Interactions among components of the *Salmonella* flagellar export apparatus and its substrates. Mol Microbiol 35(5):1052–1064. https://doi.org/10.1046/j.1365-2958.2000.01771.x

Minamino T, Imada K (2015) The bacterial flagellar motor and its structural diversity. Trends Microbiol 23(5):267–274. https://doi.org/10.1016/j.tim.2014.12.011

Minamino T, Iino T, Kutuskake K (1994) Molecular characterization of the *Salmonella typhimurium flhB* operon and its protein products. J Bacteriol 176(24):7630–7637. https://doi.org/10.1128/jb.176.24.7630-7637.1994

Minamino T, Yamaguchi S, Macnab RM (2000) Interaction between FliE and FlgB, a proximal rod component of the flagellar basal body of *Salmonella.* J Bacteriol 182(11):3029–3036. https://doi.org/10.1128/jb.182.11.3029-3036.2000

Minamino T, Kazetani KI, Tahara A, Suzuki H, Furukawa Y, Kihara M, Namba K (2006) Oligomerization of the bacterial flagellar ATPase FliI is controlled by its extreme N-terminal region. J Mol Biol 360(2):510–519. https://doi.org/10.1016/j.jmb.2006.05.010

Minamino T, Yoshimura SD, Morimoto YV, Gonzalez-Pedrajo B, Kami-Ike N, Namba K (2009) Roles of the extreme N-terminal region of FliH for efficient localization of the FliH-FliI complex to the bacterial flagellar type III export apparatus. Mol Microbiol 74(6):1471–1483. https://doi.org/10.1111/j.1365-2958.2009.06946.x

Minamino T, Imada K, Namba K (2008) Molecular motors of the bacterial flagella. Curr Opin Struct Biol 18(6):693–701. https://doi.org/10.1016/j.sbi.2008.09.006

Minamino T, Shimada M, Okabe M, Saijo-Hamano Y, Imada K, Kihara M, Namba K (2010) Role of the C-terminal cytoplasmic domain of FlhA in bacterial flagellar type III protein export. J Bacteriol 192(7):1929–1936. https://doi.org/10.1128/JB.01328-09

Minamino T, Imada K, Kinoshita M, Nakamura S, Morimoto YV, Namba K (2011a) Structural insight into the rotational switching mechanism of the bacterial flagellar motor. PLoS Biol 9(5):12. https://doi.org/10.1371/journal.pbio.1000616

Minamino T, Morimoto YV, Hara N, Namba K (2011b) An energy transduction mechanism used in bacterial flagellar type III protein export. Nat Commun 2:475. https://doi.org/10.1038/ncomms1488

Minamino T, Kinoshita M, Hara N, Takeuchi S, Hida A, Koya S, Glenwright H, Imada K, Aldridge PD, Namba K (2012) Interaction of a bacterial flagellar chaperone FlgN with FlhA is required for efficient export of its cognate substrates. Mol Microbiol 83(4):775–788. https://doi.org/10.1111/j.1365-2958.2011.07964.x

Minamino T, Kinoshita M, Inoue Y, Morimoto YV, Ihara K, Koya S, Hara N, Nishioka N, Kojima S, Homma M, Namba K (2016a) FliH and FliI ensure efficient energy coupling of flagellar type III protein export in *Salmonella.* Microbiologyopen 5(3):424–435. https://doi.org/10.1002/mbo3.340

Minamino T, Morimoto YV, Hara N, Aldridge PD, Namba K (2016b) The bacterial flagellar type III export gate complex is a dual fuel engine that can use both H^+ and Na^+ for flagellar protein export. PLoS Pathog 12(3):e1005495. https://doi.org/10.1371/journal.ppat.1005495

Minamino T, Terahara N, Kojima S, Namba K (2018) Autonomous control mechanism of stator assembly in the bacterial flagellar motor in response to changes in the environment. Mol Microbiol 109(6):723–734. https://doi.org/10.1111/mmi.14092

Minamino T, Kawamoto A, Kinoshita M, Namba K (2020a) Molecular organization and assembly of the export apparatus of flagellar type III secretion systems. Curr Top Microbiol Immunol. 427: 91–107. https://doi.org/10.1007/82_2019_170

Minamino T, Inoue Y, Kinoshita M, Namba K (2020b) FliK-driven conformational rearrangements of FlhA and FlhB are required for export switching of the flagellar protein export apparatus. J Bacteriol 202(3): e00637-19. https://doi.org/10.1128/JB.00637-19

Miyanoiri Y, Hijikata A, Nishino Y, Gohara M, Onoue Y, Kojima S, Kojima C, Shirai T, Kainosho M, Homma M (2017) Structural and functional analysis of the C-terminal region of FliG, an essential motor component of *Vibrio* Na^+-driven flagella. Structure 25(10):1540–1548, e1543. https://doi.org/10.1016/j.str.2017.08.010

Morimoto YV, Minamino T (2014) Structure and function of the bi-directional bacterial flagellar motor. Biomolecules 4(1):217–234. https://doi.org/10.3390/biom4010217

Morimoto YV, Che YS, Minamino T, Namba K (2010a) Proton-conductivity assay of plugged and unplugged MotA/B proton channel by cytoplasmic pHluorin expressed in *Salmonella*. FEBS Lett 584(6):1268–1272. https://doi.org/10.1016/j.febslet.2010.02.051

Morimoto YV, Nakamura S, Kami-ike N, Namba K, Minamino T (2010b) Charged residues in the cytoplasmic loop of MotA are required for stator assembly into the bacterial flagellar motor. Mol Microbiol 78(5):1117–1129. https://doi.org/10.1111/j.1365-2958.2010.07391.x

Morimoto YV, Nakamura S, Hiraoka KD, Namba K, Minamino T (2013) Distinct roles of highly conserved charged residues at the MotA-FliG interface in bacterial flagellar motor rotation. J Bacteriol 195(3):474–481. https://doi.org/10.1128/JB.01971-12

Morimoto YV, Ito M, Hiraoka KD, Che YS, Bai F, Kami-Ike N, Namba K, Minamino T (2014) Assembly and stoichiometry of FliF and FlhA in *Salmonella* flagellar basal body. Mol Microbiol 91(6):1214–1226. https://doi.org/10.1111/mmi.12529

Morimoto YV, Kami-Ike N, Miyata T, Kawamoto A, Kato T, Namba K, Minamino T (2016) High-resolution pH imaging of living bacterial cells to detect local pH differences. mBio 7(6). https://doi.org/10.1128/mBio.01911-16

Muramoto K, Macnab RM (1998) Deletion analysis of MotA and MotB, components of the force-generating unit in the flagellar motor of *Salmonella.* Mol Microbiol 29(5):1191–1202. https://doi.org/10.1046/j.1365-2958.1998.00998.x

Muskotal A, Seregelyes C, Sebestyen A, Vonderviszt F (2010) Structural basis for stabilization of the hypervariable D3 domain of *Salmonella* flagellin upon filament formation. J Mol Biol 403(4):607–615. https://doi.org/10.1016/j.jmb.2010.09.024

Nakamura S, Minamino T (2019) Flagella-driven motility of bacteria. Biomolecules 9(7):279. https://doi.org/10.3390/biom9070279

Nakamura S, Hanaizumi Y, Morimoto YV, Inoue Y, Erhardt M, Minamino T, Namba K (2019) Direct observation of speed fluctuations of flagellar motor rotation at extremely low load close to zero. Mol Microbiol 113 vhttps://doi.org/10.1111/mmi.14440

Namba K, Vonderviszt F (1997) Molecular architecture of bacterial flagellum. Q Rev Biophys 30(1):1–65. https://doi.org/10.1017/s0033583596003319

Okabe M, Minamino T, Imada K, Namba K, Kihara M (2009) Role of the N-terminal domain of FliI ATPase in bacterial flagellar protein export. FEBS Lett 583(4):743–748. https://doi.org/10.1016/j.febslet.2009.01.026

Paul K, Gonzalez-Bonet G, Bilwes AM, Crane BR, Blair D (2011) Architecture of the flagellar rotor. EMBO J 30(14):2962–2971. https://doi.org/10.1038/emboj.2011.188

Paul K, Erhardt M, Hirano T, Blair DF, Hughes KT (2008) Energy source of flagellar type III secretion. Nature 451(7177):489–492. https://doi.org/10.1038/nature06497

Paul K, Nieto V, Carlquist WC, Blair DF, Harshey RM (2010) The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. Mol Cell 38(1):128–139. https://doi.org/10.1016/j.molcel.2010.03.001

Postel S, Deredge D, Bonsor DA, Yu X, Diederichs K, Helmsing S, Vromen A, Friedler A, Hust M, Egelman EH, Beckett D, Wintrode PL, Sundberg EJ (2016) Bacterial flagellar capping proteins adopt diverse oligomeric states. eLife 5:e18857. https://doi.org/10.7554/eLife.18857

Pourjaberi SNS, Terahara N, Namba K, Minamino T (2017) The role of a cytoplasmic loop of MotA in load-dependent assembly and disassembly dynamics of the MotA/B stator complex in the bacterial flagellar motor. Mol Microbiol 106(4):646–658. https://doi.org/10.1111/mmi.13843

Ravid S, Eisenbach M (1984) Minimal requirements for rotation of bacterial flagella. J Bacteriol 158(3):1208–1210

Reid SW, Leake MC, Chandler JH, Lo CJ, Armitage JP, Berry RM (2006) The maximum number of torque-generating units in the flagellar motor of *Escherichia coli* is at least 11. Proc Natl Acad Sci USA 103(21):8066–8071. https://doi.org/10.1073/pnas.0509932103

Saijo-Hamano Y, Uchida N, Namba K, Oosawa K (2004) In vitro characterization of FlgB, FlgC, FlgF, FlgG, and FliE, flagellar basal body proteins of *Salmonella.* J Mol Biol 339(2):423–435. https://doi.org/10.1016/j.jmb.2004.03.070

Saijo-Hamano Y, Matsunami H, Namba K, Imada K (2019) Architecture of the Bacterial Flagellar Distal Rod and Hook of *Salmonella*. Biomolecules 9(7):260. https://doi.org/10.3390/biom9070260

Sakai T, Inoue Y, Terahara N, Namba K, Minamino T (2018) A triangular loop of domain D1 of FlgE is essential for hook assembly but not for the mechanical function. Biochem Biophys Res Commun 495(2):1789–1794. https://doi.org/10.1016/j.bbrc.2017.12.037

Sakai T, Miyata T, Terahara N, Mori K, Inoue Y, Morimoto YV, Kato T, Namba K, Minamino T (2019) Novel insights into conformational rearrangements of the bacterial flagellar switch complex. mBio 10(2):e00079–19. https://doi.org/10.1128/mBio.00079-19

Samatey FA, Imada K, Nagashima S, Vonderviszt F, Kumasaka T, Yamamoto M, Namba K (2001) Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. Nature 410(6826):331–337. https://doi.org/10.1038/35066504

Samatey FA, Matsunami H, Imada K, Nagashima S, Shaikh TR, Thomas DR, Chen JZ, Derosier DJ, Kitao A, Namba K (2004) Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. Nature 431(7012):1062–1068. https://doi.org/10.1038/nature02997

Sarkar MK, Paul K, Blair D (2010) Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. Proc Natl Acad Sci USA 107(20):9370–9375. https://doi.org/10.1073/pnas.1000935107

Shibata S, Matsunami H, Aizawa SI, Wolf M (2019) Torque transmission mechanism of the curved bacterial flagellar hook revealed by cryo-EM. Nat Struct Mol Biol 26(10):941–945. https://doi.org/10.1038/s41594-019-0301-3

Song WS, Cho SY, Hong HJ, Park SC, Yoon SI (2017) Self-oligomerizing structure of the flagellarcap protein FliD and its implication in filament assembly. J Mol Biol 429(6):847–857. https://doi.org/10.1016/j.jmb.2017.02.001

Stocker BAD (1949) Measurement of the rate of mutation of flagellar antigenic phase in *Salmonella* typhimurium. J Hyg 47:398–413

Suzuki H, Yonekura K, Namba K (2004) Structure of the rotor of the bacterial flagellar motor revealed by electron cryomicroscopy and single-particle image analysis. J Mol Biol 337(1):105–113. https://doi.org/10.1016/j.jmb.2004.01.034

Suzuki Y, Morimoto YV, Oono K, Hayashi F, Oosawa K, Kudo S, Nakamura S (2019) Effect of the MotA(M206I) mutation on torque generation and stator assembly in the *Salmonella* H^+-driven flagellar motor. J Bacteriol 201(6):e00727-e818. https://doi.org/10.1128/JB.00727-18

Takekawa N, Terahara N, Kato T, Gohara M, Mayanagi K, Hijikata A, Onoue Y, Kojima S, Shirai T, Namba K, Homma M (2016) The tetrameric MotA complex as the core of the flagellar motor stator from hyperthermophilic bacterium. Sci Rep 6:31526. https://doi.org/10.1038/srep31526

Terahara N, Kodera N, Uchihashi T, Ando T, Namba K, Minamino T (2017a) Na^+-induced structural transition of MotPS for stator assembly of the *Bacillus* flagellar motor. Sci Adv 3(11):eaao4119. https://doi.org/10.1126/sciadv.aao4119

Terahara N, Noguchi Y, Nakamura S, Kami-Ike N, Ito M, Namba K, Minamino T (2017b) Load- and polysaccharide-dependent activation of the Na^+-type MotPS stator in the *Bacillus subtilis* flagellar motor. Sci Rep 7:46081. https://doi.org/10.1038/srep46081

Terahara N, Inoue Y, Kodera N, Morimoto YV, Uchihashi T, Imada K, Ando T, Namba K, Minamino T (2018) Insight into structural remodeling of the FlhA ring responsible for bacterial flagellar type III protein export. Sci Adv 4(4):eaao7054. https://doi.org/10.1126/sciadv.aao7054

Thomas J, Stafford GP, Hughes C (2004) Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export. Proc Natl Acad Sci USA 101(11):3945–3950. https://doi.org/10.1073/pnas.0307223101

Tipping MJ, Delalez NJ, Lim R, Berry RM, Armitage JP (2013) Load-dependent assembly of the bacterial flagellar motor. mBio 4(4). https://doi.org/10.1128/mBio.00551-13

Ueno T, Oosawa K, Aizawa S (1992) M ring, S ring and proximal rod of the flagellar basal body of Salmonella typhimurium are composed of subunits of a single protein, FliF. J Mol Biol 227(3):672–677. https://doi.org/10.1016/0022-2836(92)90216-7

Vartanian AS, Paz A, Fortgang EA, Abramson J, Dahlquist FW (2012) Structure of flagellar motor proteins in complex allows for insights into motor structure and switching. J Biol Chem 287(43):35779–35783. https://doi.org/10.1074/jbc.C112.378380

Yamaguchi S, Aizawa S, Kihara M, Isomura M, Jones CJ, Macnab RM (1986) Genetic evidence for a switching and energy-transducing complex in the flagellar motor of *Salmonella typhimurium*. J Bacteriol 168(3):1172–1179. https://doi.org/10.1128/jb.168.3.1172-1179.1986

Yamaguchi T, Toma S, Terahara N, Miyata T, Ashihara M, Minamino T, Namba K, Kato T (2020) Structural and functional comparison of *Salmonella* flagellar filaments composed of FljB and FliC. Biomolecules 10(2):246. https://doi.org/10.3390/biom10020246

Yonekura K, Maki S, Morgan DG, DeRosier DJ, Vonderviszt F, Imada K, Namba K (2000) The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. Science 290(5499):2148–2152. https://doi.org/10.1126/science.290.5499.2148

Yonekura K, Maki-Yonekura S, Namba K (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature 424(6949):643–650. https://doi.org/10.1038/nature01830

Yuan J, Berg HC (2008) Resurrection of the flagellar rotary motor near zero load. Proc Natl Acad Sci USA 105(4):1182–1185. https://doi.org/10.1073/pnas.0711539105

Zhou J, Fazzio RT, Blair DF (1995) Membrane topology of the MotA protein of *Escherichia coli*. J Mol Biol 251(2):237–242. https://doi.org/10.1006/jmbi.1995.0431

Zhou J, Lloyd SA, Blair DF (1998a) Electrostatic interactions between rotor and stator in the bacterial flagellar motor. Proc Natl Acad Sci USA 95(11):6436–6441. https://doi.org/10.1073/pnas.95.11.6436

Zhou J, Sharp LL, Tang HL, Lloyd SA, Billings S, Braun TF, Blair DF (1998b) Function of protonatable residues in the flagellar motor of *Escherichia coli*: a critical role for Asp 32 of MotB. J Bacteriol 180(10):2729–2735

Chapter 9
Cellulosomes: Highly Efficient Cellulolytic Complexes

Victor D. Alves, Carlos M. G. A. Fontes, and Pedro Bule

Abstract Cellulosomes are elaborate multienzyme complexes capable of efficiently deconstructing lignocellulosic substrates, produced by cellulolytic anaerobic microorganisms, colonizing a large variety of ecological niches. These macromolecular structures have a modular architecture and are composed of two main elements: the cohesin-bearing scaffoldins, which are non-catalytic structural proteins, and the various dockerin-bearing enzymes that tenaciously bind to the scaffoldins. Cellulosome assembly is mediated by strong and highly specific interactions between the cohesin modules, present in the scaffoldins, and the dockerin modules, present in the catalytic units. Cellulosomal architecture and composition varies between species and can even change within the same organism. These differences seem to be largely influenced by external factors, including the nature of the available carbon-source. Even though cellulosome producing organisms are relatively few, the development of new genomic and proteomic technologies has allowed the identification of cellulosomal components in many archea, bacteria and even some primitive eukaryotes. This reflects the importance of this cellulolytic strategy and suggests that cohesin-dockerin interactions could be involved in other non-cellulolytic processes. Due to their building-block nature and highly cellulolytic capabilities, cellulosomes hold many potential biotechnological applications, such as the conversion of lignocellulosic biomass in the production of biofuels or the development of affinity based technologies.

Keywords Cellulosome · Modular · CAZymes · Cohesin-Dockerin · Scaffoldin · Biofuels

V. D. Alves · C. M. G. A. Fontes · P. Bule (✉)
CIISA, Faculdade de Medicina Veterinária, ULisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal
e-mail: pedrobule@fmv.ulisboa.pt

V. D. Alves
e-mail: vdalves@fmv.ulisboa.pt

C. M. G. A. Fontes
e-mail: cafontes@fmv.ulisboa.pt

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_9

The Cellulosome—Introduction

Plant cell wall polysaccharides are one of the most abundant sources of carbon and energy on Earth. They mainly consist of cellulose fibers, which are usually associated with hemicellulose, lignin (a non-saccharide polymer), pectins and proteins. The interactions between the various plant cell wall components form an intricate structure that is highly resistant to degradation, making all the stored energy largely inaccessible (Albersheim et al. 2011; Carpita et al. 2015). Because of this complexity, plant cell wall polysaccharides require a large number of highly specialized enzymes to catalyse their degradation. These enzymes are included in a group of biocatalysts capable of degrading, modifying or creating glycosidic bonds, termed Carbohydrate Active Enzymes or CAZymes (Henrissat and Davies 2000; Lombard et al. 2010). In light of its potential as a clean alternative source of renewable energy, the conversion of lignocellulosic biomass into fermentable sugars has fuelled a significant amount of dedicated research, leading to the identification and characterization of an ever expanding number of CAZymes (Jørgensen et al. 2007; Cantarel et al. 2009; Lombard et al. 2014). Most of these enzymes are produced by cellulolytic microorganisms, including bacteria, fungi and protozoa, which are able to hydrolyse the lignocellulose in plant cell walls, recycling the photosynthetically fixed carbon and therefore playing a crucial role in the carbon cycle (Himmel and Bayer 2009; Himmel et al. 2010). In general, aerobic cellulolytic species secret copious amounts of enzymes which act individually during structural polysaccharide hydrolysis. The released products are then captured by the cellulolytic cell and used as an energy source. However, the energetic constraints posed by anaerobic ecosystems limit the enzyme producing capabilities of microorganisms. This has imposed an evolutionary pressure on anaerobic bacteria and fungi leading to the development of an alternative and more efficient cellulolytic system, termed the cellulosome (Bayer et al. 1994; Fontes and Gilbert 2010).

The cellulosome was first described in the early 1980's, in the gram-positive thermophilic bacteria *Hungateiclostridium thermocellum* (*Clostridium thermocellum*) (Bayer et al. 1983; Lamed et al. 1983). This macromolecular structure consists of a remarkably efficient and well-organized cell surface multi-enzyme complex that allows enzyme recycling and direct assimilation of polysaccharide hydrolysis products. Cellulosomes have several different components, but are essentially made of two types of building blocks: dockerin (Doc) bearing enzymes and cohesin (Coh) bearing structural proteins, named scaffoldins (Bayer et al. 2004; Fontes and Gilbert 2010; Artzi et al. 2017). The high affinity and high specificity protein-protein interactions established between cohesin and dockerin modules define the molecular platform for cellulosome assembly. These two components bind tightly with each other according to a specificity pattern, dictating final cellulosomal architecture and composition (Fig. 9.1) (Carvalho et al. 2003; Doi and Kosugi 2004; Hamberg et al. 2014).

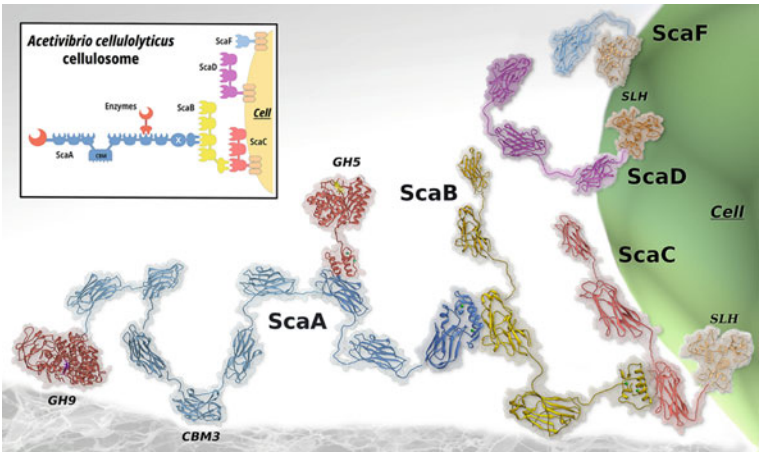


Fig. 9.1 Acetivibrio cellulolyticum cellulosome. A structure-based concept art illustration. The cellulosome architecture is organized around the non-catalytic primary scaffoldin (ScaA), a unique adaptor scaffoldin (ScaB) and several anchoring scaffoldins (ScaC, ScaD and ScaF). ScaA bears 9 cohesin (Coh) modules that congregate the catalytic enzymes *via* their appended dockerin (Doc) modules, it also can bind the carbohydrate substrate through a carbohydrate binding module (CBM) and it is crowned by a N-terminal glycoside hydrolase (GH) and by a C-terminal X-Doc that enables the anchoring of the cellulosome to the bacterial cell-wall. The schematic cellulosome in the small image insert is color-coded according to the main 3D image representation. The quaternary structure as well as the unstructured linkers between modules, are an "artistic" rendering speculation of a cellulosome. The structure artwork was generated in UCSF Chimera v 1.15 *Adapted* from (Bule et al. 2018b)

In this chapter, we dissect the cellulosome in order to explore the functional and structural properties of each of its key components. We also focus on cellulosomal diversity by discussing some of the known cellulosome-producing organisms, including the architecture and composition of their complexes. Finally, we look at the regulation of cellulosomal gene transcription and discuss some of the potential biotechnological applications for these macromolecular structures.

Cellulosomal Components

Cellulosomes result from the assembly of several distinct proteins, usually displaying a modular organization, with each module having a specific function within the macromolecular cellulolytic complex. Those functions include cellulosome assembly, structural support, catalytic activity, substrate binding and cellular tethering. The assembly elements are what distinguish the cellulosomal system from the free enzyme systems. They include the cohesin modules, present in purely structural non-catalytic proteins, termed scaffoldins, and the dockerin modules, present

in the enzymes and some scaffoldins. The strong and specific protein-protein interactions occurring between cohesin and dockerin modules allow the several cellulosomal components to congregate in a highly-efficient cellulolytic nanomachine. Nonetheless, several other elements are also present in the cellulosome, contributing to its catalytic efficiency, flexibility and ability to recognize its substrates. Here, we break down the cellulosome into its fundamental pieces in order to provide a better understanding of their importance within the macromolecular complex.

The Scaffoldin

The backbone that creates the cellulosome is built around scaffoldins, which are structural proteins mainly composed of linker spaced multiple cohesin modules (Fig. 9.2). According to their function, scaffoldins are classified into four major types: primary, anchoring, adaptor and cell-free scaffoldins. It is around primary scaffoldins that cellulosomes are assembled, by simultaneously incorporating up to hundreds of CAZymes through high affinity Coh-Doc interactions (Artzi et al. 2017; Smith et al. 2017). Cohesins and dockerins are classified into types (I, II, III, R) based on primary structure homology (Bayer et al. 2004). Apart from a few exceptions, type I interactions are usually committed to cellulosome assembly, while type II interactions are associated with cell wall tethering. Therefore, most known primary scaffoldins possess a C-terminal type II dockerin, which interacts with the complementary type II cohesin modules in cell-bound anchoring scaffoldins. Many scaffoldins have an Ig-like stabilizing module, termed the X-module, adjacent to the N-terminal end of the type II dockerin (Adams et al. 2006). Another common feature in primary scaffoldins is the presence of a non-catalytic domain with carbohydrate-binding activity. These modules, named Carbohydrate-Binding Modules (CBMs), target the cellulosome to the crystalline regions of the plant cell wall. Interestingly, scaffoldin CBMs are almost exclusively family 3 CBMs, as it will be further discussed below (Tormo et al. 1996; Shimon et al. 2000; Hong et al. 2014).

Complex cellulosomes have dedicated scaffoldins responsible for mediating the attachment of the cellulosome to the bacterial cell surface, called anchoring scaffoldins (Fig. 9.2). The importance of this cell-bound arrangement *vis-a-vis* the alternative cell-free cellulosome is reflected by the clustering of cellulosomal genes, with primary and anchoring scaffoldin genes usually clustered together (Zhivin et al. 2017). Anchoring scaffoldins are non-covalently bound to the organism surface through an SLH domain, or covalently bound through a sortase recognition motif (Lemaire et al. 1995; Rincon et al. 2005). Another type of scaffoldin, the adaptor scaffoldin, has also been described on more elaborate and highly structured cellulosomes (Fig. 9.2). Polyvalent adaptor scaffoldins that have multiple cohesin modules, can serve as an amplifier for the number of CAZymes integrated into the cellulosome (Xu et al. 2003; Dassa et al. 2012). Alternatively, monovalent adaptor scaffoldins use a single cohesin module to intermediate binding between enzymes bearing divergent

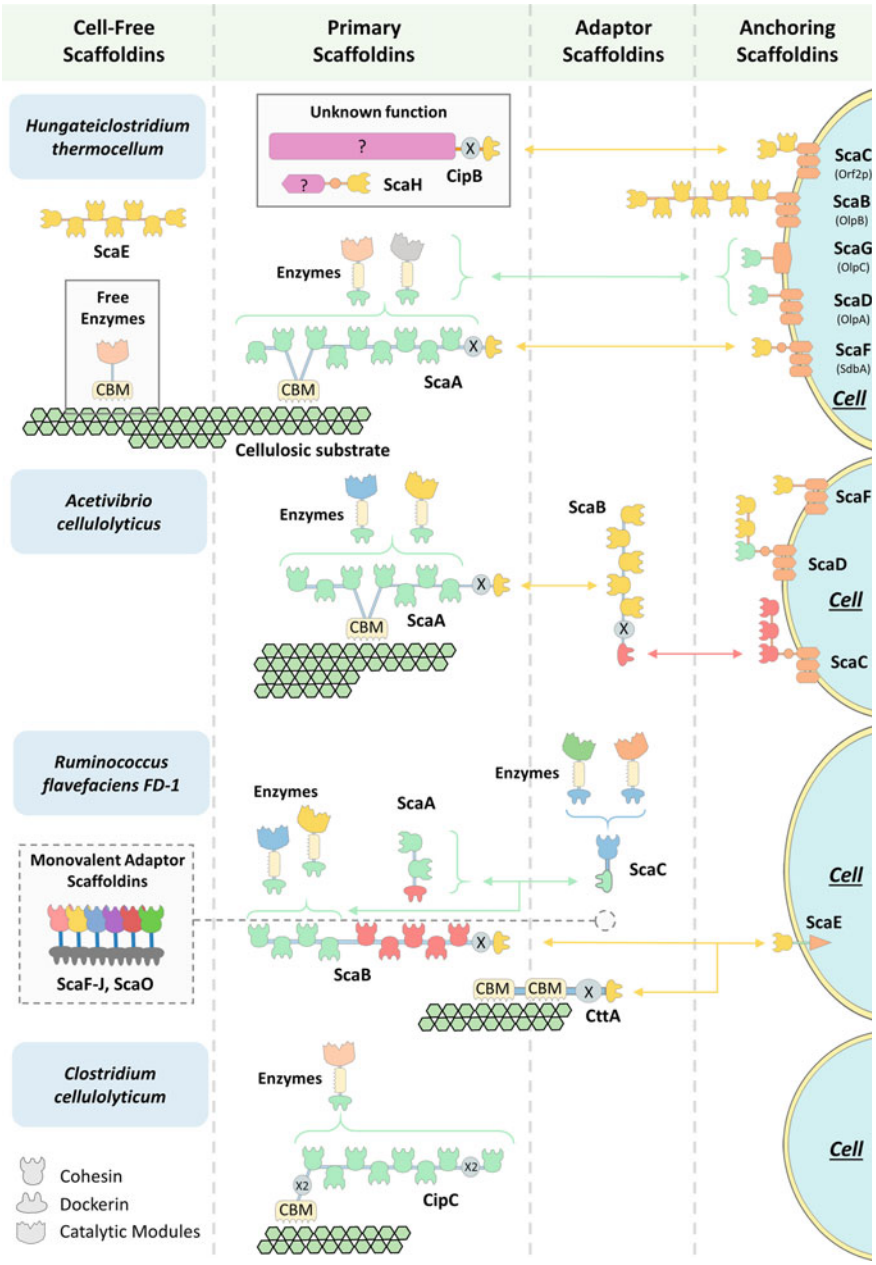


Fig. 9.2 Architecture of cellulosome systems. Schematic representation of four distinct cellulosomal systems belonging to the bacteria *H. thermocellum*, *A. cellulolyticus*, *R. flavefaciens* and *C. cellulolyticus*. Within each system, the same color is used for dockerins sharing similar specificity and for the cohesins recognized by those dockerins. X-modules are represented by gray circles marked with an X. Polyssacharides are represented as green hexagons

dockerins and the primary scaffoldin (Fig. 9.2). They can also function as a repertoire change inducer, by switching the specificity of the CAZymes that populate the primary scaffoldin in response to substrate availability (Rincon et al. 2004; Moraïs et al. 2016a). Cell-free scaffoldins are exclusively composed of cohesin modules, separated by linker sections (Fig. 9.2). They originate untethered cellulosomes that are released to the extracellular space, like the ones secreted by *Clostridium clarifavum* or *Hungateiclostridium thermocellum*, which can coexist with cell-bound cellulosomes (Artzi et al. 2015; Xu et al. 2016).

Some noteworthy exceptions from this general scaffoldin blueprint are the absence of a C-terminal dockerin in the primary scaffoldins of simple cellulosomes (Dassa et al. 2017) and the reversed cohesin-dockerin arrangement of *Bacteroides (Pseudobacteroides) cellulosolvens* cellulosome system. In this species, the assembly of enzymes onto the primary scaffoldin relies on type II cohesins, while the terminal type I dockerin mediates cellulosome attachment to an anchoring scaffoldin (Zhivin et al. 2017).

Cohesins and Dockerins

Cohesin and dockerin modules are undoubtedly the most important components of the cellulosome, such that the basis for the identification of cellulosomal proteins is dependent upon the presence of one or both of these modular elements. These complementary modules are generally classified based on sequence similarity. In the archetypal cellulosome of *Hungateiclostridium thermocellum*, type I interactions are reserved for CAZYme recruitment onto the primary scaffoldin, while type II interactions allow tethering the cellulosome to the bacterial cell surface. Although other species present similar arragements, this dichotomy was broken upon the characterization of the highly complex cellulosome of *Ruminococcus flavefaciens* (Berg Miller et al. 2009), which possesses divergent cohesins and dockerins, subsequently classified as type III modules. More recently, a comprehensive bioinformatics analysis of *Bacteroides (Pseudobacteroides) cellulosolvens*'s genome, led to the discovery of yet another type of interaction, named type R, in what is currently the most complex cellulosome system known to date (Zhivin et al. 2017). As more data are generated from genomic and metagenomics studies, it seems that the type classification for cohesins and dockerins might be more diverse than initially suspected. Metagenomic data mining offers new opportunities when dealing with uncultured bacteria and such analysis have already uncovered some divergent unclassified cohesins that do not fall in any of the known types (Bensoussan et al. 2017). Structural homology type classification versus a functional classification, although convenient, can be misleading, and a number of noteworthy discrepancies imply a cautious approach, namely the already mentioned type I and II function reversal of *Bacteroides (Pseudobacteroides) cellulosolvens* (Xu et al. 2004a; Bule et al. 2018b).

Cohesin-dockerin interactions are one of the strongest non-covalent protein-protein interactions known (Stahl et al. 2012) and the basis for cellulosomal assembly.

They tend to be species-specific, although cross-specificity (predominantly for type II interactions) has been previously described (Pagès et al. 1997; Haimovitz et al. 2008; Fontes and Gilbert 2010; Bule et al. 2018b). Both cohesins and dockerins display a striking overall conservation of structural topology. Cohesins are approximately 150 amino acids long, arranged in a flattened and elongated β-sandwich, composed of nine β-strands divided between two β-sheets (A and B), arranged in a jellyroll topology, and bearing a hydrophobic core. β-sheet A is composed of β-strands 4, 7, 2, 1 and 9, while β-sheet B is composed of β-strands 5, 6, 3 and 8 (Fig. 9.3a, b). A distinctive and noticeable planar plateau is formed by β-sheet B which establishes the dockerin binding interface of cohesins (Carvalho et al. 2003; Adams et al. 2006). Structurally, dockerins are short protein modules (~ 70 residues long) with an α-helix dominated configuration, arranged around two highly conserved duplicate sequences (N-terminal or helix-1 and C-terminal or helix-3) presenting nearly perfect 2-fold symmetry, each composed of about 22 amino acid that, alternatively, comprise the cohesin binding sites of the dockerin. These two tandem repeats are connected by a smaller third α-helix (helix-2), and hold a stable tertiary configuration based upon two conserved F-hand Ca^{2+}-binding loop-helix structural motifs (Fig. 9.3). An almost fully conserved Gly residue on the N-term helix-1 constitutes a null reference (#0) for a convenient relative dockerin residue position numbering. As a consequence, residues at positions #1, 3, 5, 9, and 12 (and the equivalent positions on the C-term helix-2) are typically aspartate or asparagine residues and are involved in calcium-coordination, whereas positions #10 and 11 are recognized as main residues involved in cohesin recognition, along with positions #17, 18 and 22. A prevailing residue conservation on both helix-1 and -3 cohesin recognition sites and the discovery that some dockerin can bind to their cohesin partners by either an N- or C-terminal helix dominated interaction, disclosed the existence of a dual-binding mode for type-I interactions (Fig. 9.3) (Carvalho et al. 2003). This means that some dockerins can bind to the same partner in two distinct orientations, which are related by a 180º rotation relatively to the cohesin. On the other hand, the first structurally characterized type II cohesin-dockerin complex revealed a single-binding mode of interaction. In this complex, the lack of residue conservation between the two dockerin repeats only allows binding in a single orientation, with no dominating helix, since residues on both helices are significantly involved in the interaction. Interestingly, in type II complexes, the adjacent X-module also participates in cohesin recognition (Adams et al. 2006; Brás et al. 2012; Salama-Alber et al. 2013).

Initially, it was believed that the dual-binding mode would be ubiquitous for cellulosome CAZyme recruitment, contributing to avert steric clashes among cellulosome components while facilitating substrate access, by providing two alternative binding conformations. In contrast, the single-binding mode would be the norm for cellulosome cell-anchoring. However, it is now clear that in some organisms, like *Acetivibrio cellulolyticus*, there is a widespread dual-binding mode interaction in both cellulosome assembly and anchoring, even when it involves adaptor scaffoldins (Cameron et al. 2015; Brás et al. 2016; Bule et al. 2018a). Furthermore, ruminococal cellulosomes, like that of *Ruminococcus flavefacien*s, can be assembled and anchored exclusively through single binding mode cohesin-dockerin interactions (Bule et al.

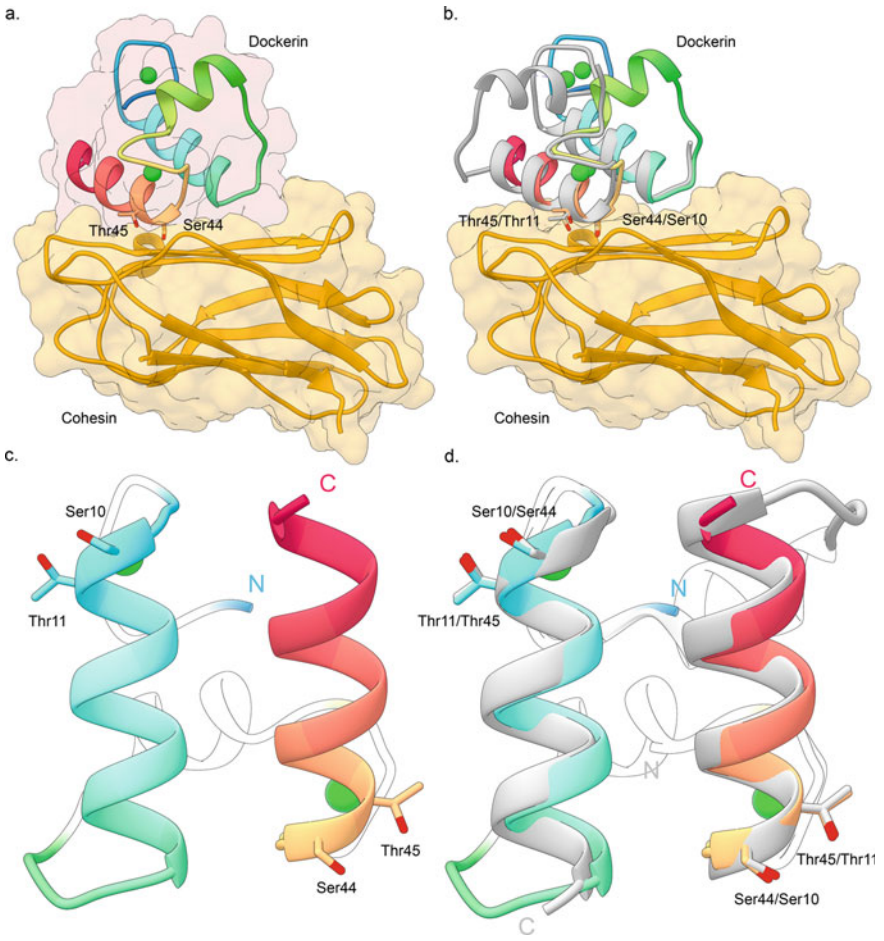


Fig. 9.3 The dual binding mode. **a**. Ribbon representation of a type I cohesin-dockerin complex from *Hungateiclostridium thermocellum* (PDB accession code 1OHZ), containing the dockerin from a xylanase (in rainbow coloring: N-terminus is blue, C-terminus is red) and the second cohesin from primary scaffoldin ScaA (in yellow). In this panel it is possible to observe that the dockerin is interacting with the cohesin mainly through its C-terminal helix. **b**. Overlay of 1OHZ with the same Coh-Doc complex (PDB accession code 2CCL) in which the dockerin (in grey) has been mutated to force interaction with the cohesin through its N-terminal helix. The overlay was obtained by matching the cohesins of both complexes. It is possible to observe that the dockerin of 2CCL is rotated 180° relative to the dockerin in 1OHZ. **c**. Cohesin binding interface of the type I dockerin present in the 1OHZ Coh-Doc complex (N-terminus in blue, C-terminus in red). **d**. Comparison of the two cohesin binding helices of the type I Doc, by overlaying the dockerin with a version of itself rotated by 180° (in grey). This comparison shows that there is conservation of the two key cohesin recognition residues (Serine/Threonine pair) at the same relative positions (#10, #11) in both helices. Van der Waals' surfaces are represented in translucent colors and calcium ions are shown in green. The structures in this figure were generated using UCSF Chimera v 1.13.1

2016, 2017). Interestingly, the lack of conservation between the two symmetric dockerin halves has, in some cases, resulted in a dual-specificity-binding strategy, in which each of the helices at the binding interface recognizes a different cohesin. This means that the dockerin can bind distinct partners by using opposite orientations, revealing the enormous versatility of such interactions (Voronov-Goldman et al. 2015; Brás et al. 2016). Ultimately, such diversity, although based on a very conserved main chain Cα trace, relies on the particular composition of key interface recognition amino acid residues, defining its conserved or divergent character, and subtly modifying inter- and intra-species barriers, binding specificity and binding mode for the cohesin-dockerin interaction (Slutzki et al. 2015; Bule et al. 2016, 2017; Brás et al. 2016).

Enzymes

Cellulosomal enzymes are, by definition, enzymes containing a dockerin domain. These enzymes are typically modular and, in addition to the dockerin that mediates protein-protein interactions, possess one or more catalytic modules and sometimes one or more CBMs, involved in protein-carbohydrate interactions (Fontes and Gilbert 2010). As a plant cell wall degrading macromolecular complex, the cellulosome can integrate numerous different carbohydrate active enzymes. The first identified cellulosomal enzymes were cellulases, which is not surprising considering that the first cellulosome was discovered due to its highly efficient cellulolytic properties (Lamed et al. 1983). In addition to cellulases, other cellulosomal enzymes were subsequently identified, most notably xylanases, pectinases, carbohydrate esterases, mannanases and xyloglucanases.

Much like their substrates, complex polysaccharide degrading enzymes are extremely diverse and tend to be highly specific. They include glycoside hydrolases, carbohydrate esterases and polysaccharide lyases and have been grouped into families according to primary sequence homology, kept in an ever-growing online database called CAZy.org (Henrissat and Davies 1997, 2000; Cantarel et al. 2009; Lombard et al. 2010, 2014). The degradation of complex polysaccharides often requires the synergistic action of several different CAZymes. Synergy may be due to different modes of action towards the same substrate, like when an endoglucanase hydrolyses the substrate, thereby producing additional chain ends that can be cleaved by an exoglucanase. It can also result from the hydrolysis of two tightly associated substrates, in which the action of one enzyme could make the concealed substrate accessible for the action of the second enzyme (e.g. cellulases and xylanases). It can also work by limiting product inhibition effects, when a second enzyme acts on the product of a primary enzyme, thus restoring the activity of the first enzyme (e.g. β-glucosidases and cellulases) (Morag et al. 1991). Even the hydrolysis of a relatively simple polysaccharide such as cellulose involves the cooperative efforts of at least three groups of enzymes: endo-(1,4)-β-D-glucanases, exo-(1,4)-β-D-glucanases and β-glucosidases. The endoglucanase randomly attacks the internal O-glycosidic

bonds, resulting in glucan chains of different lengths; the exoglucanase acts on the ends of the cellulose chains and releases β-cellobiose as the major end product and finally, the β-glucosidases act specifically on the β-cellobiose disaccharides releasing glucose (Kuhad et al. 2011).

The genome sequencing of several cellulosome producing bacteria has identified a large variety of dockerin bearing enzymes, which include not only cellulases but also other CAZymes such as hemicellulases, ligninases (Fan and Yuan 2010; Cragg et al. 2015), pectinases (Tamaru and Doi 2001; Chakraborty et al. 2015), mannanases (Kurokawa et al. 2001; Sabathé et al. 2002; Perret et al. 2004) and chitinases, suggesting a complex mechanism of enzymatic synergy. Nonetheless, almost all known cellulosome-producing bacteria characteristically express large amounts of a single family 48 glycoside hydrolase (GH48) exoglucanase, which is a key component for the highly efficient cellulolytic activity displayed by these structures (Morag et al. 1991; Ravachol et al. 2015). Contrastingly, an extensive repertoire of family 9 glycoside hydrolases is generally secreted by these bacteria, which work in close synergy with GH48 since, on its own, GH48 displays very little activity on cellulose. Many GH9 enzymes from different cellulosomic systems have been characterized, displaying different activities, distinct abilities to bind to cellulosic substrates and diverse synergies with the major GH48 exoglucanase (Ravachol et al. 2014; Moraïs et al. 2016b). This eventually suggests that enzyme diversity, especially of GH9 enzymes, reflects the structural diversity of the substrates. In fact, the nature of the available substrate seems to be an important factor in cellulosome composition and efficiency, as revealed by some proteomic studies (Gold and Martin 2007; Cho et al. 2010; Tsai et al. 2010; Wang et al. 2019). The expression levels of enzymatic components can change depending on the carbon source, which means that cellulosomes with different compositions can be assembled by the same microorganism when grown in the presence of different substrates. Other glycoside hydrolase families, such as GH5, GH10, GH11 and GH43, are also commonly found in cellulosome systems, meaning that these complexes are able to utilize a large variety of plant cell wall polysaccharides (Artzi et al. 2017).

Interestingly, dockerin bearing proteins are not restricted to CAZymes. Many other activities have been reported for potential cellulosomic proteins, which are not directly linked to polyssacharide degradation. These include serpins (Kang et al. 2006), proteases (Levy-Assaraf et al. 2013) and expansins (Chen et al. 2016; Artzi et al. 2017). The role of these biocatalysts in cellulosome function is still unclear although it has been suggested that they may have protective and regulatory functions or even act in synergy with CAZymes to degrade highly complex substrates (Chen et al. 2016; Moraïs et al. 2016b; Bensoussan et al. 2017).

Carbohydrate Binding Modules

In general, multi-modular CAZymes are appended to non-catalytic Carbohydrate Binding Modules (CBMs) that bind carbohydrates, thus playing a critical role in

targeting their associated enzymes towards the substrate, resulting in improved catalytic activity (Boraston et al. 2004; Gilbert et al. 2013). Since the first described CBMs had crystalline cellulose as their primary ligand, these non-catalytic polysaccharide recognizing modules were initially defined as Cellulose-Binding Domains, CBDs (Gilkes et al. 1988). After these initial studies, it has been shown that many carbohydrates can be targeted by different 'CBDs', including crystalline cellulose, non-crystalline cellulose, chitin, β-1,3-glucans, β-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan and starch (Boraston et al. 2004). Some can even display 'lectin-like' specificity and bind to a variety of cell-surface glycans (Fujimoto et al. 2013). In order to reflect their diverse ligand specificity, these modules were renamed Carbohydrate Binding Modules, or CBMs (Boraston et al. 1999).

Much like the catalytic modules of CAZymes, CBMs are classified into families in the CAZy database, according to primary structure homology. Currently there are 86 different CBM families registered on cazy.org, a number that has been consistently growing (Boraston et al. 2004; Lombard et al. 2014). According to their function, CBMs have also been classified into one of three types: type A CBMs which interact with crystalline polysaccharides (primarily cellulose); type B modules which bind to internal regions of single glycan chains, and type C CBMs that recognize small saccharides in the context of a complex carbohydrate (Boraston et al. 2004; Fontes and Gilbert 2010; Armenta et al. 2017).

It is now clearly recognized that CBMs potentiate CAZyme function (Bolam et al. 1998). However, the mechanism by which they do so is still not fully understood. It was suggested that CBMs may function through a proximity effect or have a more active role leading to a disruptive effect on the polysaccharides (Boraston et al. 2004; Guillén et al. 2009). Many studies have shown that maintaining enzymes in the proximity of the insoluble substrate leads to a more efficient degradation, as the removal of CBMs from enzymes results in a dramatic reduction in enzymatic activity (Bolam et al. 1998; Boraston et al. 2003). Curiously, enzyme activity on soluble substrates is not frequently affected when CBMs are removed (Kleine and Liebl 2006), meaning that CBM function is highly dependent on the nature of the substrate.

CBMs can also be found in many cellulosomal systems, usually occupying a central position within the primary scaffoldin (Boraston et al. 2004; Fontes and Gilbert 2010). Generally, primary scaffoldins contain a family 3 CBM, such as the one present in ScaA of *H. thermocellum* (Poole 1992; Morag et al. 1995). This is a type A CBM that binds strongly to the crystalline surface of cellulose, accounting for the primary targeting of the cellulosome onto its substrate (Bayer et al. 2004; Artzi et al. 2017). In contrast to what has been described for many species such as *H. thermocellum*, *A. cellulolyticus* or *P. cellulosolvens*, the scaffoldins produced by ruminal bacterium *R. flavefaciens* do not possess a CBM module targeting cellulosic substrates. Rather, this bacterium expresses an independent 'CttA' scaffoldin with two putative CBMs and a dockerin that specifically recognizes an anchoring scaffoldin, which may serve to attach the bacterial cell directly to the substrate (Fig. 9.2) (Rincon et al. 2010; Salama-Alber et al. 2013). Interestingly, a cell-free scaffoldin produced by *A. cellulolyticus*, was found to contain two family 2 CBMs as opposed

to a family 3 CBM (Dassa et al. 2012). Comparable CBM2-containing cell-free scaffoldins were also observed in the genome of *Clostridium clariflavum* (Artzi et al. 2014)(Artzi et al. 2014). These are the only known examples of scaffoldins containing CBMs from a family other than CBM3. Family 2 CBMs are more commonly found associated with free, non-cellulosomal enzymes and the fact that they have only been found in cell-free scaffoldins suggests that they could have a distinct function in cellulosomes to that of family 3 CBMs (Simpson et al. 2000; Artzi et al. 2017).

CBMs seem to play a vital role in cellulosome function and not only as a part of the scaffoldin proteins. Although only two CBM families have been reported in scaffoldins, many cellulosomal enzymes have CBMs from various families, exhibiting different carbohydrate specificities. Presently, the study of cellulosomes has been an important source of novel carbohydrate binding molecules and has significantly helped our understanding of the mechanisms behind carbohydrate recognition (Venditto et al. 2016; Ribeiro et al. 2019). Much like free CAZymes, the enzyme-borne cellulosomal CBMs likely work to position the catalytic modules in the proximity of the substrate, resulting in optimal hydrolysis (Hammel et al. 2005).

Anchoring Modules

As mentioned above, most cellulosome producing species assemble different cellulosomes presenting distinct conformations and, depending on the nature of their main scaffoldin, cellulosomes can either be found in a cell-bound or cell-free form (Fig. 9.2). Nonetheless, in most species there appears to be a preference for the former. From an energy-efficiency point of view, one could infer the obvious advantages that a tethered complex presents to the anaerobic microorganism since (1) there is close proximity between the products of hydrolysis and the cell, resulting in optimal product uptake and (2) the ability to retain the complex at the cell surface allows the microorganism to benefit from its activity for long periods of time, as opposed to having to express multiple copies as they diffuse away. Furthermore, using ex vivo assembled mini-cellulosomes with identical catalytic components, it has been observed that the cell-bound versions display up to 4.5 times higher hydrolytic activity than the cell-free versions. This suggests that tethering the cellulosome could also be important for its potential biotechnological applications, as it optimizes function (You et al. 2012).

The surface layer, or S-layer, is a common feature of many gram-positive bacteria. They are monomolecular crystalline arrays of proteinaceous subunits that surround the exterior cell wall, providing various functions, including protection and cellular interaction (Sleytr et al. 1993; Doi and Kosugi 2004). Since the S-layer was first described, much attention was given to finding how its proteins target and bind the cell wall. Curiously, the first clue came from the sequence of a gene cluster encoding some of *H. thermocellum* scaffoldins, including ScaA (then called CipA). The cluster sequence revealed an ORF encoding a cohesin module and a C-terminal domain displaying significant similarity to the N-terminal region of S-layer proteins

from various species. The authors suggested that this protein could be somehow involved in the cell-anchoring of *H. thermocellum*'s cellulosome (Fujino et al. 1993) and later confirmed its ability to bind the cell envelope (Lemaire et al. 1995). Since then, the now called SLH (S-layer homology) domain has been identified in many bacterial surface proteins. It is composed of either one or three repeated GPTS (glycine, proline, threonine, serine)-rich segments that bind peptidoglycan associated polysaccharides in a non-covalent interaction (Mesnage 2000), tethering extracellular proteins to the bacterial wall, including many cellulosome anchoring scaffoldins (Xu et al. 2004b; Zhivin et al. 2017). On the other hand, most mesophilic cellulosomes, such as the ones produced by *R. flavefaciens* and *Ruminococcus champanelensis*, do not possess SLH domains in their anchoring scaffoldins. Instead, they are covalently bound to the cell wall via a sortase-mediated attachment mechanism (Rincon et al. 2005; Salama-Alber et al. 2013). This mechanism involves three crucial features, present at the C-terminus of the protein, and collectively called the cell wall sorting signal: an LPXTG (leucine, proline, X, threonine and glycine, where X is any amino acid) motif, a hydrophobic domain which interacts with the cytoplasmic membrane and a tail of charged residues (Schneewind et al. 1992; Mazmanian 1999). During secretion, the hydrophobic domain and charged tail prevents protein secretion by stopping membrane translocation, allowing recognition of the LPXTG motif by the sortase enzyme. The sortase then cleaves the LPXTG motif between the threonine and glycine residues and covalently attaches the threonine to the amino group of the pentaglycine cell wall cross-bridge resulting in a cell wall attached protein (Paterson and Mitchell 2004).

The X-Module

A common feature of primary scaffoldins is the presence of an enigmatic module at the N-terminus of their dockerins, which has been termed the X module (Fig. 9.2). This X module-dockerin pairing (XDoc) has been described in many cellulosomes, such as the ones assembled by *H. thermocellum*, *A. cellulolyticus, P. cellulosolvens* and *R. flavefaciens.* The exact function of this domain is still relatively unknown, although it has been repeatedly suggested that its presence could provide structural stability and enhanced solubility to the adjacent cellulosomal components (Mosbah et al. 2000; Schubot et al. 2004; Xu et al. 2004a; Kataeva et al. 2004; Adams et al. 2005). The crystal structure of a *H. thermocellum* type II cohesin-dockerin complex together with its neighbouring X-module shows direct hydrogen-bond contacts between the Ig-like X module and the cohesin. Together with the fact that the X module stabilizes the type II dockerin in solution, this suggests that there is a functional enhancement of adjacent components in the presence of the X module (Adams et al. 2006). The authors also report an increased affinity of the type II Coh for the XDoc modular pair *versus* that for the isolated type II Doc, supporting this assumption.

Unlike the X-module in the type II Coh-XDoc interaction of *H. thermocellum*, the X-module in the type III Coh-XDoc complex of *R. flavefaciens* does not appear to contribute directly to the CohE-Doc binding surface (Fig. 9.2). Rather, its elongated stalk-like conformation, supported by 3 dockerin inserts that act as structural buttresses, appears to serve as an extended spacer, projecting the cellulosome away from the bacterial cell wall (Salama-Alber et al. 2013). This would position the catalytic domains away from the cell surface facilitating the action of the cellulosomal enzymes on the plant cell polysaccharides. In fact, the *R. flavefaciens* X module does not bear homology with any other X modules from cellulolytic bacteria. Instead, it shares significant similarity with the G5 module of *Streptococcus pneumoniae's StrH* protein, whose function is to position the catalytic modules away from the cell, allowing optimal processing of host cell surface-presented glycoconjugates.

Cellulosome Diversity

Since the cellulosome was first described, it has become increasingly apparent that its prevalence in nature is not as pronounced as initially predicted. As far as our knowledge extends, this system is rather restricted to a few specialized microorganisms inhabiting specific ecological niches such as the soil, sewage sludge and some mammalian gastrointestinal compartments. Nonetheless, there is no denying that cellulosomes play an essential role in the conversion of recalcitrant lignocellulosic substrates in anaerobic environments, and with the expanding availability of genomic, metagenomic and proteomic tools, it is very likely that new cellulosome producing species will continue to emerge.

Depending on whether cellulosomes possess a single scaffoldin or multiple types of interacting scaffoldins, cellulosome systems are divided into two major types: the simple and the complex. As mentioned, scaffoldins are the molecular platforms that organize cellulosomes and have the potential to be exceptionally varied in size and constitution, leading to cellulosome systems that are structurally diverse. Cellulosomes can be rather simple and small, like those of *Clostridium acetobutylicum*, which can only assemble 5 enzymes in its single scaffoldin (Sabathé et al. 2002). They can also be large and complex such as those of the *P. cellulosolvens* system, with its 31 different scaffoldins and a theoretical assembly that can hold up to 110 enzymes in one single cellulosome unit (Zhivin et al. 2017). As opposed to complex cellulosomes, which are found in species living in a wide range of environmental conditions, simple cellulosomes have only been identified in mesophilic organisms. These bacteria produce cellulosomes with a single scaffoldin backbone that can attach up to 9 cohesins, depending on the species, and are usually not tethered to the bacterial wall. These include *Clostridium josui* (Ichikawa et al. 2014), *Clostridium cellulovorans* (Cho et al. 2010), *Clostridium papyrosolvens* (Pohlschröder et al. 1994), *Clostridium cellulolyticum* (Fig. 9.2) (Xu et al. 2015) and *Ruminococcus bromii* (Ze et al. 2015). This last species is a key resistant-starch degrader found in the human gut (Ze et al.

2012) and, as such, its dockerin-bearing enzymes were found to hydrolyze starch, rather than cellulose. Hence, *R. bromii* complexes were referred to as amylosomes.

The first described cellulosome and currently best-characterized system, is that of the thermophilic bacterium *H. thermocellum* (Lamed et al. 1983). Interestingly, *H. thermocellum*'s highly efficient cellulolytic performance is supported by three systems: a free enzyme system; a cell-free cellulosome system and a cell-bound cellulosome system (Xu et al. 2016), which has been extensively studied and still serves as the archetypal model to this day (Fig. 9.2). *H. thermocellum*'s genome encodes 72 dockerin-containing proteins, suggesting an elaborate cellulosome system (Pinheiro et al. 2009). This bacterium can produce eight different scaffoldins including primary scaffoldin ScaA, the five cell-anchoring scaffoldins ScaB, C, D, F and G, the two cell-free scaffoldins ScaE and H and the elusive CipB. Primary scaffoldin ScaA has eight type I cohesins, separated by linkers of various lengths, which can bind to the dockerins present on cellulosomal enzymes (Gerwig et al. 1993; Kruus et al. 1995). It also possesses an internal family 3 CBM and a C-terminal type II X-dockerin that can bind to the cohesins on anchoring scaffoldins ScaB, C or F, tethering the cellulosome to the bacterial surface, or to the cohesins on ScaE, originating cell-free cellulosomes (Xu et al. 2016). Considering that ScaB and ScaE have seven type II cohesins each, it is possible for *H. thermocellum* to assemble both cell-free and cell-bound cellulosomes with up to 63 enzymes on a single unit. Enzymes can also be directly attached to the bacterial surface by interacting with the cohesins on ScaG and ScaD which, like the other anchoring scaffoldins on this bacterium, are bound to the cell wall by their C-terminal SLH domains (Leibovitz and Béguin 1996. The roles of CipB and ScaH are still not fully understood.

The concept of adaptor scaffoldin was first introduced by the cellulosome of *Acetivibrio cellulolyticus* (Fig. 9.2) which, with 16 different scaffoldins and over 140 dockerin-bearing proteins encoded in its genome, has a considerably more intricate system than that of *H. thermocellum* (Xu et al. 2003; Dassa et al. 2012). These adaptor scaffoldins promote the incorporation of more enzymes into one single complex allowing *A. cellulolyticus* to integrate up to 84 enzymes in its cellulosome, which, including ScaA's own catalytic module, totalizes 96 catalytic components. The thermophilic *Clostridium clariflavum* has a similar system to that of *A. cellulolyticus* in the sense that it also relies on adaptor scaffoldins to create larger assemblies, like its theoretical, record breaking, 160-enzyme cellulosome (Artzi et al. 2014). Interestingly, in the cellulosomal system of the ruminal bacterium *R. flavefaciens* (Fig. 9.2), monovalent adaptor scaffoldins possessing a single cohesin have also been identified. These adaptors, rather than increasing the number of enzymes, increase the repertoire of different catalytic units that can integrate the complex (Rincón et al. 2004; Bule et al. 2016). The genome of *R. flavefaciens* encodes an enormous number of cellulosomal components, including 222 dockerin-containing proteins, the largest number described so far (Rincon et al. 2010). These dockerins have been classified into six different families according to their primary structure homology, a classification that translates into function, as members of the same family recognize similar cohesin partners (Israeli-Ruimy et al. 2017; Bule et al. 2017, 2018c). The monovalent adaptor scaffoldins are capable of recruiting specific families into the

cellulosome, integrating different catalytic activities in the complex's profile (Bule et al. 2016). Intriguingly, numerous dockerin bearing proteins were also identified in the *Ruminococcus albus* genome, another cellulolytic ruminal bacterium, but only a single cohesin was found (Dassa et al. 2014). Given that *R. albus* shares the same ecological niche as *R. flavefaciens*, it has been proposed that these bacteria have a symbiotic relationship, in which the enzymes produced by *R. albus* can be integrated into *R. flavefaciens* cellulosome, with both benefiting from its catalytic action. This hypothesis is supported by the fact that the dockerins from the two species share a high degree of homology, which in some cases can be as high as 60% identity. The recently described cellulosome system of *Ruminococcus champanelensis*, a cellulolytic bacteria found in the human gut, was also found to possess a very elaborate architecture, similar to that of *R. flavefaciens* (Cann et al. 2016; Moraïs et al. 2016a).

The existence of fungal multi-enzyme complexes has also been proposed from as early as 1992 (Wilson and Wood 1992). Although it has been suggested that these complexes are assembled through cohesin-dockerin interactions (Fanutti et al. 1995), there is still very little information regarding fungal cellulosomes when compared to their bacterial counterparts. Fungal dockerins have been identified in numerous cellulolytic enzymes but differ significantly from bacterial dockerins (Haitjema et al. 2017). They can be found either at the N or C-terminus of enzymes and exist in tandem repeats rather than single copies at the C-terminal region (Steenbakkers et al. 2001). This tandem repetition seems to have a functional significance as it has been observed that a double-dockerin can bind more tightly to the cellulosome than a single domain one (Nagy et al. 2007). The only reported 3-dimensional structures of a fungal dockerin were obtained by NMR and do not display the typical double F-hand calcium binding fold seen in all bacterial dockerins, but rather a single β-sheet composed of three or four small antiparallel strands and a short α-helix (Raghothama et al. 2001; Nagy et al. 2007). Furthermore, there seems to exist pronounced interspecies recognition, with dockerins from one fungal organism binding the cellulosome of others (Nagy et al. 2007). Until recently, there was no convincing evidence supporting the existence of fungal cohesin domains, but a comparative genomics study allowed the definitive identification of scaffoldin proteins in the anaerobic fungi *Anaeromyces robustus*, *Neocallimastix californiae* and *Piromyces finnis* (Haitjema et al. 2017). These scaffoldin proteins were shown to bind non-catalytic dockerin domains present in enzymes and contained multiple repeating motifs, which were suggested to be the cohesin domains. The minimum sequence that defines a single fungal cohesin remains to be determined but scaffoldin fragments with only four of these repeats were capable of dockerin recognition. Clearly, there is still much left to know about fungal cellulosomes but they could potentially have distinct advantages over the bacterial ones such as a larger variety of catalytic activities and the fact that they produce glucose and not cellobiose as final product (Gilmore et al. 2015). It is becoming increasingly evident that fungi have a key role in the conversion of recalcitrant biomass, much due to their cellulosomal enzymes (Hagen et al. 2020). This makes fungal cellulosomes attractive blueprints for the engineering of complexes with multiple biotechnological applications (Gilmore et al. 2020).

Recent advances in genome-sequencing methods and new metagenomic approaches are promoting the rapid discovery of new cellulosome systems and expanding our understanding of well-known cellulosomal species (Dassa et al. 2012, 2014; Artzi et al. 2014; Zhivin et al. 2017). Cohesins and dockerins are being described in the genome of uncultured bacteria within species not previously documented (Bensoussan et al. 2017). New types of cohesins that do not fit within the previously described three types were also found, suggesting that these modules are more diverse than originally predicted. In addition, cohesins and dockerins have been identified in the genomes of non-cellulolytic species such as *Clostridium perfringens*, an opportunistic pathogen that inhabits the gastrointestinal tract (Bayer et al. 1999; Adams et al. 2008; Peer et al. 2009). Some of these dockerin-bearing proteins were found to be toxins or to form multi-protein complexes that promote virulence. This could mean that the cohesin-dockerin interaction can be present in other systems besides the cellulosome, possibly serving other purposes other than polysaccharide degradation.

Quaternary Structure

Even though there is now abundant information regarding the 3-dimensional structures of each of the several cellulosomal components, very little is known about the conformation of an entire complex. After the first successful attempt at determining the structure of a cohesin-dockerin complex (Carvalho et al. 2003), some were inspired to crystallize larger portions of cellulosomes, albeit with relatively little success (Currie et al. 2012). The intrinsic plasticity of linker segments connecting the various cohesin modules, allow scaffoldins to adopt a variety of dramatically different conformations, making cellulosome quaternary architecture extremely hard to decipher. This conformational flexibility between cohesin modules allow for the optimal positioning of the enzymatic subunits onto the substrate (Hammel et al. 2005). In fact, data comparing the activity of designer cellulosomes containing different linker lengths between cohesins suggest that longer linkers improve hydrolytic efficiency (Vazana et al. 2013). Interestingly, the linker section connecting the catalytic cores of cellulosomal enzymes with their dockerins seem to be predominantly rigid (Hammel et al. 2004, 2005).

The largest crystal structure of a cellulosome ever to be determined by X-ray crystallography belongs to the canonical cellulosome of *H. thermocellum* and includes five modules from three different proteins (Currie et al. 2012). It contains the type I Doc of Cel9D bound to a C-terminal trimodular fragment of ScaA scaffoldin, which includes a type I cohesin connected by a linker to the X module-type II dockerin dyad, which in turn is interacting with ScaF's type II cohesin (Fig. 9.4). The structure reveals an elongated topology with a flexible 13-residue linker connecting the ninth type I cohesin (CohI9) module and the X module. Elevated temperature factors suggest that the linker is highly dynamic. In addition, the four copies of the penta-modular complex found in the asymmetric unit of the crystal structure had

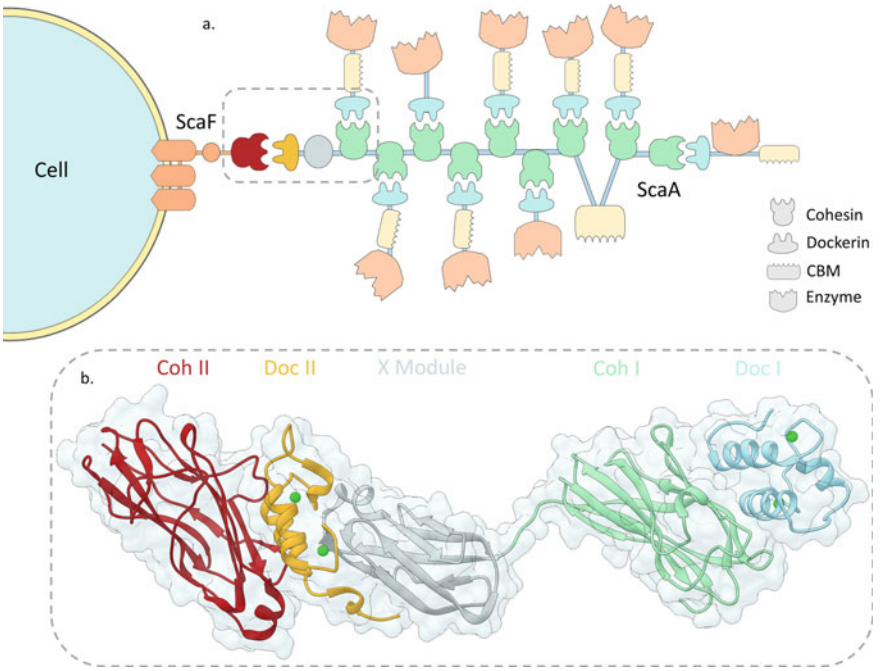


Fig. 9.4 Tridimensional structure of the largest crystallized cellulosome fragment. **a**. Cellulosome of *Hungateiclostridium thermocellum* resulting from the interaction of anchoring scaffoldin ScaF and primary scaffoldin ScaA. The modules within the dashed box correspond to the modules seen in the structure below. **b**. Ribbon representation of a ternary cellulosome complex structure from *Hungateiclostridium thermocellum,* containing a total of five modules (PDB accession code 4FL4). The type II cohesin from anchoring scaffoldin ScaF is represented in red and is interacting with the type II dockerin from primary scaffoldin ScaA, represented in yellow. The ScaA fragment also shows the dockerin adjacent X-module (in grey) separated from the 9th type II cohesin (in green) by a linker region. Lastly, a type I dockerin from the enzyme Cel9D can be seen in light blue interacting with ScaA's 9th cohesin. The van der Waals' surface of the complex is shown in translucent blue. Calcium ions are represented in green. The structure in this figure was generated using UCSF Chimera v 1.13.1

slightly different orientations in the DocI-CohI9 region. This suggests the existence of several possible conformations of the linker sequences that contribute to modulate the overall conformation of the cellulosome.

Regulation of Cellulosomal Components

The genes encoding most cellulosomal subunits are organized in clusters on the bacterial chromosome. On simpler systems, the scaffoldin gene is followed downstream by the dockerin-bearing enzyme encoding genes, while complex systems have

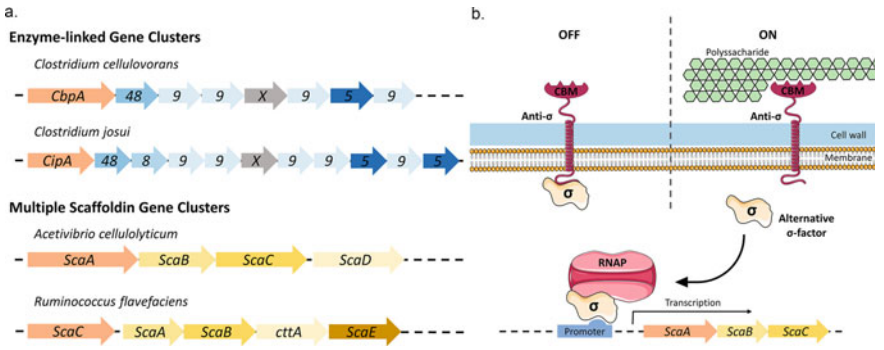


Fig. 9.5 Transcription regulation of cellulosomal genes. **a**. Examples of gene clusters present in the genome of two simple cellulosome producing bacteria and two complex cellulosome producing bacteria. **b**. Transcription regulation via a substrate sensitive anti sigma factor and alternative sigma factor mechanism. The CBM-like extracellular domain of the anti-sigma factor (in purple) recognizes the appropriate polyssacharide (green hexagons) and changes conformation releasing the alternative sigma factor (**σ**) inside the bacterial cell. The alternative sigma factor binds to the RNA polymerase (RNAP) which will then be able to recognize the appropriate promoter and initiate transcription of the cellulosomal gene. This figure was made using Servier Medical ART (SMART)

multiple scaffoldin clusters (Fig. 9.5a). As seen in the supra mentioned examples, the repertoire of dockerin bearing enzymes encoded by each cellulosome producing system far exceeds the number of cohesins available, even on some of the largest assemblies. For example, *R. flavefaciens* FD-1 can gather up to 14 enzymes on a single cellulosome while its genome encodes over 220 dockerin containing proteins (Rincon et al. 2010). Even *B. cellulosolvens,* whose largest assembly can theoretically hold up to 110 enzymes, possesses over 200 dockerin-bearing ORFs in its genome (Zhivin et al. 2017). One of the few species whose cellulosome could eventually attach its full repertoire of dockerin-bearing proteins (78 in total) simultaneously on a single unit is *C. clariflavum*, with its theoretical 160 cohesin-bearing assembly (Artzi et al. 2014). This suggests that cellulosome composition is not preestablished and is probably influenced by external factors. Growth rate seems to influence the expression of cellulosomal genes, with high growth rates resulting in down-regulation of the components (Dror et al. 2003a). More importantly though, is that the expression of these genes can change according to the nature of the available carbon source, resulting in a diversity of possible cellulosomal configurations within the same system (Dror et al. 2003b; Han et al. 2003; Raman et al. 2009; Artzi et al. 2015). It was previously observed that the expression levels of cellulases, such as the highly prevalent GH48 exoglucanase, are increased when the bacteria are grown on cellulose compared to when cultivated on cellobiose, while the opposite is true for other glycoside hydrolases. Enzymatic profiles also seem to differ in the presence of hemicellulosic polysaccharides when compared to those of cultures grown on crystalline cellulose (Dykstra et al. 2014; Artzi et al. 2015).

It is evident that a substrate sensing mechanism is operating within cellulosome producing organisms, in order to adjust the expression of cellulosomal components to

the available carbon sources. Some of the bacteria producing complex cellulosomes, such as *H. thermocellum*, *C. clariflavum*, *A. cellulolyticus* and *B. cellulosolvens*, seem to regulate their cluster genes by a pathway involving a set of transmembrane and RNA polymerase interacting proteins called anti-sigma factors and alternative sigma factors, respectively (Kahel-Raifer et al. 2010; Nataf et al. 2010; Muñoz-Gutiérrez et al. 2016). This is a common mechanism found in bacteria, in which the sigma factor (or alternative sigma factor) interacts with an RNA polymerase, promoting gene transcription, while the anti-sigma factor suppresses transcription by sequestering the sigma factor (Paget 2015). The ingenuity of the cellulosome regulating mechanism is in the extracellular portion of the anti-sigma factor involved, which has CBM-like domains capable of recognizing specific polysaccharides in the surrounding medium. When a transmembrane anti-sigma factor binds to its cognate substrate it changes conformation, releasing the alternative sigma factor in the intracellular space, which will in turn interact with the RNA polymerase, allowing promoter recognition and initiating transcription (Fig. 9.5b). This allows the bacteria to optimize its resources, by recognizing the polysaccharides in the extracellular medium and reacting accordingly, changing cellulosome composition and architecture as new substrates are exposed during the degradation process. A substrate independent transcriptional regulation mechanism has also been reported in *C. cellulolyticum*, which regulates cellulosomal stoichiometry by selective RNA processing and stabilization (Xu et al. 2015). This would allow the production of pre-optimized cellulosomes which could be crucial for correct cellulolytic function.

Applications for Cellulosomes

Due to its building-block nature, it is no surprise that the cellulosome has inspired numerous ingenious biotechnological solutions. These mainly explore the highly specific cohesin-dockerin interaction to promote reaction mechanisms whose components function in close proximity. The development of mini-cellulosomes and designer cellulosomes represent one of the most popular and successful studies in the field of cellulosomics (Bayer et al. 1994). The in vitro and in vivo construction of these cellulosome derivatives have been used both in the study of cellulosomal function and in trying to improve the efficiency of native cellulosomes, for various biotechnological purposes. The main difference between mini-cellulosomes and designer cellulosomes lies in the nature of their cohesins. The first are built on truncated versions of the native scaffoldins, containing cohesins similar to those found in the original complexes. Designer cellulosomes, on the other hand, have a chimeric scaffoldin composed of cohesins with distinct specificities that can originate from multiple species, allowing the incorporation of specific enzymes with control over position and number of copies (Fierobe et al. 2002, 2005; Moraïs et al. 2012). By designing tailor-made scaffoldins, several authors have been able to answer numerous questions regarding cellulosome assembly and function. These include the dynamics of enzyme incorporation into scaffoldins (Borne et al. 2013), the effects

of enzyme relative position (Stern et al. 2015), the importance of scaffoldin-borne CBMs for catalytic efficiency (Mingardon et al. 2007) and the importance of intermodular linkers for flexibility and steric accommodation (Caspi et al. 2009; Molinier et al. 2011; Vazana et al. 2013), among other findings.

One of the most extensively explored applications of designer cellulosomes has been its use in the conversion of lignocellulosic biomass into fermentable sugar, which is still a major bottleneck in biofuel production (Hasunuma et al. 2013). For this purpose, cellulosomes are used as a blueprint for the design, construction and exploitation of tailor-made catalytic multi-protein complexes integrating fungal and bacterial enzymes from non-aggregating systems, displaying particular promise in biomass saccharification. These enzymes include some not commonly found in cellulosomes, such as lytic polysaccharide monooxygenases (LPMOs), expansins and laccases, which have successfully helped to improve cellulosome efficiency with the introduction of novel complementary enzymatic activities (Gefen et al. 2012; Arfi et al. 2014; Chen et al. 2016; Davidi et al. 2016). Notably, there have been some efforts to obtain genetically engineered microorganisms, capable of producing and displaying mini-cellulosomes or designer cellulosomes on their cell surface. The objective is to create bacteria suitable for consolidated bioprocessing (CBP), meaning that they can be used to ferment plant cell walls into fuels in a one step process, using a single bioreactor, without the addition of enzymes (Hyeon et al. 2011; Moraïs et al. 2014; Ou and Cao 2014; Liang et al. 2014; Willson et al. 2016).

Other innovative applications have been exploring the properties of cellulosomal components, mainly taking advantage of the ability for CBMs to recognize polysaccharides and the high-affinity and specificity of cohesin-dockerin interactions. These include microarray technologies (Haimovitz et al. 2008; Israeli-Ruimy et al. 2017), affinity protein-purification systems (Demishtein et al. 2010) and molecular biosensors (Hyeon et al. 2014). To surpass the limitation posed by the near irreversible cohesin-dockerin interaction, a *H. thermocellum* dockerin with a lower affinity for its cohesin partner was successfully engineered, which enabled its use as an affinity tag for protein purification (Demishtein et al. 2010). The protein purification system consists of a *H. thermocellum* CBM fused to a cohesin that can be immobilized in a cellulose column and is capable of recognizing the dockerin tag in a recombinant protein. The protein is then eluted with EDTA, which disrupts the cohesin-dockerin interaction. Besides proving a very efficient and robust alternative system for affinity chromatography, the affinity tag was also shown to have little effect on the properties of the proteins tested, including enzymes. Furthermore, the relatively inexpensive costs of cellulose-based affinity columns together with their reusable nature and high capacity make this system very attractive for affinity protein purification (Karpol et al. 2009; Demishtein et al. 2010). Another curious cohesin-dockerin based technology consists of a high sensitivity and selectivity self-assembling biosensor that could potentially be used to detect multiple specific molecules in a complex mixture, such as blood serum (Hyeon et al. 2014).

It has been shown that the proximity effect promoted by cellulosomes can also be transferred to other platforms. By drawing inspiration from cellulosomal assembly, other cellulolytic macromolecular structures have been designed. These include

self-assembling circular multi-enzyme complexes (Mitsuzawa et al. 2009), cellulase coated nanospheres (Blanchette et al. 2012), streptavidin clustered cellulases (Kim et al. 2012) and cellulases that are bound to DNA scaffolds (Mori et al. 2013). Nevertheless, none of these platforms seem to possess the same degree of structural flexibility or substrate targeting capabilities which accounts for much of the cellulosome's efficiency.

The utilisation of enzymes and cellulosomes are also being considered as valuable alternatives for the usage of agro-wastes and organic pollutants as a renewable resource, reducing the consequent environmental pollution (Bayer et al. 2007; Karmakar and Ray 2011) and as feed additives to improve digestibility of animal diets (Costa et al. 2014). Even so, only a few of the many possible research applications have been explored and, as our knowledge about these multi-enzyme complexes increases, so does the potential for future innovation.

Concluding Remarks

There is no doubt that the study of cellulosomes has provided considerable insights on how recalcitrant polysaccharides are processed in nature, highlighting the pivotal role that anaerobic microorganisms colonizing lignocellulosic environments have in the carbon cycle. In addition, deciphering cellulosome function led to the discovery of many new biocatalysts and carbohydrate interacting molecules (Venditto et al. 2016) with many potential biotechnological applications, with the conversion of cellulosic biomass into biofuels being the most explored (Bayer et al. 2010; Arora et al. 2015; Kahn et al. 2019). The strong and highly-specific interaction between cohesin and dockerin modules has also broaden our knowledge about protein-protein interactions and provided innovative tools for the development of new affinity based technologies, with applications in research and healthcare (Haimovitz et al. 2008; Demishtein et al. 2010; Hyeon et al. 2014).

There is still much to be explored in the field of cellulosomics. Although cellulosome producing microorganisms are relatively rare, new systems will no doubt continue to emerge at increasingly faster rates. The massification of next generation sequencing technologies, metagenomic approaches, high-throughput methodologies and automation have streamlined otherwise very time-consuming protocols and are allowing the identification of cellulosomal components in a large variety of organisms, including some non-cellulolytic species (Adams et al. 2008; Peer et al. 2009; Dassa et al. 2014; Ben David et al. 2015; Haitjema et al. 2017). It would be extremely interesting to understand how wide cellulosome diversity can extend and what other functions could the cohesin-dockerin interaction potentially have in nature, beyond cellulosomal assembly. Regulation of cellulosomal component expression and particularly its dependence on substrate availability has been somewhat elucidated through transcriptomic and proteomic studies, but there is still a lack of information regarding their secretion and the assembly processes (Blouzard et al. 2010; Kahel-Raifer et al. 2010). Fungal cellulosomes are also still very poorly understood, especially regarding

the nature of the cohesin-dockerin interactions supporting their assembly. Clearly defined cohesin modules are yet to be described and whether glycosylation is important for dockerin binding remains to be established (Nagy et al. 2007). Given the extensive repertoire of dockerin-bearing fungal enzymes, understanding the fungal cellulosome system is an objective worth pursuing, as it could provide answers to many of the challenges found in biomass conversion (Haitjema et al. 2017).

One particularly challenging task that has been eluding researchers for many years is the determination of a high-resolution structure of a complete cellulosome. Such structure could shed light into the interactions between the several cellulosomal components and between the cellulosome and the substrate. It could also provide the answer to whether the dual-binding mode actually serves to better accommodate the enzymes in the complex, avoiding steric hindrance. Unfortunately, the highly flexible nature of cellulosomes and the heterogeneity in enzyme content make them impractical crystallization subjects. Cryo-electron microscopy could prove useful to that end, but for now only low-resolution small-angle X-ray scattering structures are available (Hammel et al. 2005; Currie et al. 2013).

The development of novel designer cellulosomes is one of the most prolific strategies in the field of cellulosomics. Besides providing several tools for the study of cellulosomes, many studies have been developing new complexes with improved biochemical properties destined for biotechnological purposes (Arora et al. 2015; Kahn et al. 2019; Gilmore et al. 2020), including the production of biofuels by CBP. It is to be expected that new cellulosome based technology will continue to emerge, in not only the field of biofuel production, but also in those where affinity-based approaches should prove useful. To that end, it is essential to further our knowledge about the mechanics behind these intricate systems.

References

Adams JJ, Webb BA, Spencer HL, Smith SP (2005) Structural characterization of type II dockerin module from the cellulosome of Clostridium thermocellum: calcium-induced effects on conformation and target recognition. Biochemistry 44:2173–2182. https://doi.org/10.1021/bi048039u

Adams JJ, Pal G, Jia Z, Smith SP (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. Proc Natl Acad Sci USA 103:305–310. https://doi.org/10.1073/pnas.0507109103

Adams JJ, Gregg K, Bayer EA et al (2008) Structural basis of Clostridium perfringens toxin complex formation. Proc Nat Acad Sci 105:12194–12199. https://doi.org/10.1073/pnas.0803154105

Albersheim P, Darvill A, Roberts K et al (2011) Plant cell walls: from chemistry to biology. Garland Science

Arfi Y, Shamshoum M, Rogachev I et al (2014) Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. Proc Natl Acad Sci USA 111:9109–9114. https://doi.org/10.1073/pnas.1404148111

Armenta S, Moreno-Mendieta S, Sánchez-Cuapio Z et al (2017) Advances in molecular engineering of carbohydrate-binding modules: engineering of carbohydrate-binding modules. Proteins: Struct, Funct, Bioinf 85:1602–1617. https://doi.org/10.1002/prot.25327

Arora R, Behera S, Sharma NK, Kumar S (2015) Bioprospecting thermostable cellulosomes for efficient biofuel production from lignocellulosic biomass. Bioresources and Bioprocessing 2. https://doi.org/10.1186/s40643-015-0066-4

Artzi L, Dassa B, Borovok I et al (2014) Cellulosomics of the cellulolytic thermophile Clostridium clariflavum. Biotechnol Biofuels 7:100. https://doi.org/10.1186/1754-6834-7-100

Artzi L, Morag E, Barak Y et al (2015) Clostridium clariflavum: key cellulosome players are revealed by proteomic analysis. MBio 6:e00411–e00415. https://doi.org/10.1128/mBio.00411-15

Artzi L, Bayer EA, Moraïs S (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol 15:83–95. https://doi.org/10.1038/nrmicro.2016.164

Bayer EA, Kenig R, Lamed R (1983) Adherence of Clostridium thermocellum to cellulose. J Bacteriol 156:818–827

Bayer EA, Morag E, Lamed R (1994) The cellulosome–a treasure-trove for biotechnology. Trends Biotechnol 12:379–386. https://doi.org/10.1016/0167-7799(94)90039-6

Bayer EA, Coutinho PM, Henrissat B (1999) Cellulosome-like sequences in archaeoglobus fulgidus: an enigmatic vestige of cohesin and dockerin domains. FEBS Lett 463:277–280

Bayer EA, Belaich J-P, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554. https://doi.org/10.1146/annurev.micro.57.030502.091022

Bayer EA, Lamed R, Himmel ME (2007) The potential of cellulases and cellulosomes for cellulosic waste management. Curr Opin Biotechnol 18:237–245. https://doi.org/10.1016/j.copbio.2007.04.004

Bayer EA, Lamed R, White BA et al (2010) Conversion of agricultural residues to bioethanol: the roles of cellulases and cellulosomes. In: Blaschek HP, Ezeji TC, Scheffran J (eds) Biofuels from agricultural wastes and byproducts. Wiley-Blackwell, Oxford, UK, pp 67–96

Ben David Y, Dassa B, Borovok I et al (2015) Ruminococcal cellulosome systems from rumen to human. Environ Microbiol 17:3407–3426. https://doi.org/10.1111/1462-2920.12868

Bensoussan L, Moraïs S, Dassa B et al (2017) Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. Environ Microbiol 19:185–197. https://doi.org/10.1111/1462-2920.13561

Berg Miller ME, Antonopoulos DA, Rincon MT et al (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of ruminococcus flavefaciens FD-1. PLoS ONE 4:e6650. https://doi.org/10.1371/journal.pone.0006650

Blanchette C, Lacayo CI, Fischer NO et al (2012) Enhanced cellulose degradation using cellulase-nanosphere complexes. PLoS ONE 7:e42116. https://doi.org/10.1371/journal.pone.0042116

Blouzard J-C, Coutinho PM, Fierobe H-P et al (2010) Modulation of cellulosome composition in Clostridium cellulolyticum: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. Proteomics 10:541–554. https://doi.org/10.1002/pmic.200900311

Bolam DN, Ciruela A, McQueen-Mason S et al (1998) Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. Biochem J 331(Pt 3):775–781

Boraston AB, McLean BW, Kormos JM et al (1999) Carbohydrate-binding modules: diversity of structure and function. In: Gilbert HJ, Davies GJ, Henrissat B, Svensson B (eds) Recent advances in carbohydrate bioengineering. Royal Society of Chemistry, Cambridge, pp 202–211

Boraston AB, Kwan E, Chiu P et al (2003) Recognition and hydrolysis of noncrystalline cellulose. J Biol Chem 278:6120–6127. https://doi.org/10.1074/jbc.M209554200

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781. https://doi.org/10.1042/bj20040892

Borne R, Bayer EA, Pagès S et al (2013) Unraveling enzyme discrimination during cellulosome assembly independent of cohesin-dockerin affinity. FEBS J 280:5764–5779. https://doi.org/10.1111/febs.12497

Brás JLA, Alves VD, Carvalho AL et al (2012) Novel Clostridium thermocellum type I cohesin-dockerin complexes reveal a single binding mode. J Biol Chem 287:44394–44405. https://doi.org/10.1074/jbc.M112.407700

Brás JLA, Pinheiro BA, Cameron K et al (2016) Diverse specificity of cellulosome attachment to the bacterial cell surface. Sci Rep 6:38292. https://doi.org/10.1038/srep38292

Bule P, Alves VD, Leitão A, et al (2016) Single-binding mode integration of hemicellulose degrading enzymes via adaptor scaffoldins in Ruminococcus flavefaciens cellulosome. J Biol Chem jbc.M116.761643. https://doi.org/10.1074/jbc.M116.761643

Bule P, Alves VD, Israeli-Ruimy V et al (2017) Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes. Sci Rep 7:759. https://doi.org/10.1038/s41598-017-00919-w

Bule P, Cameron K, Prates JAM et al (2018a) Structure–function analyses generate novel specificities to assemble the components of multi-enzyme bacterial cellulosome complexes. J Biol Chem jbc.RA117.001241. https://doi.org/10.1074/jbc.RA117.001241

Bule P, Pires VM, Fontes CM, Alves VD (2018b) Cellulosome assembly: paradigms are meant to be broken! Curr Opin Struct Biol 49:154–161. https://doi.org/10.1016/j.sbi.2018.03.012

Bule P, Pires VMR, Alves VD, et al (2018c) Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction. Scientific Reports 8:. https://doi.org/10.1038/s41598-018-25171-8

Cameron K, Najmudin S, Alves VD et al (2015) Cell-surface attachment of bacterial multienzyme complexes involves highly dynamic protein-protein anchors. J Biol Chem 290:13578–13590. https://doi.org/10.1074/jbc.M114.633339

Cann I, Bernardi RC, Mackie RI (2016) Cellulose degradation in the human gut: Ruminococcus champanellensis expands the cellulosome paradigm. Environ Microbiol 18:307–310. https://doi.org/10.1111/1462-2920.13152

Cantarel BL, Coutinho PM, Rancurel C et al (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkn663

Carpita N, Ralph J, McCann M (2015) Cell walls. In: Buchanan B, Gruissem W, Jones R (eds) Biochemistry & molecular biology of plants, 2nd edn. Wiley Blackwell, Oxford, UK, pp 45–110

Carvalho AL, Dias FMV, Prates JAM et al (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. Proc Natl Acad Sci USA 100:13809–13814. https://doi.org/10.1073/pnas.1936124100

Caspi J, Barak Y, Haimovitz R et al (2009) Effect of linker length and dockerin position on conversion of a thermobifida fusca endoglucanase to the cellulosomal mode. Appl Environ Microbiol 75:7335–7342. https://doi.org/10.1128/AEM.01241-09

Chakraborty S, Fernandes VO, Dias FMV et al (2015) Role of pectinolytic enzymes identified in Clostridium thermocellum cellulosome. PLoS ONE 10:e0116787. https://doi.org/10.1371/journal.pone.0116787

Chen C, Cui Z, Song X et al (2016) Integration of bacterial expansin-like proteins into cellulosome promotes the cellulose degradation. Appl Microbiol Biotechnol 100:2203–2212. https://doi.org/10.1007/s00253-015-7071-6

Cho W, Jeon SD, Shim HJ et al (2010) Cellulosomic profiling produced by Clostridium cellulovorans during growth on different carbon sources explored by the cohesin marker. J Biotechnol 145:233–239. https://doi.org/10.1016/j.jbiotec.2009.11.020

Costa M, Fernandes VO, Ribeiro T et al (2014) Construction of GH16 β-glucanase mini-cellulosomes to improve the nutritive value of barley-based diets for broilers. J Agric Food Chem 62:7496–7506. https://doi.org/10.1021/jf502157y

Cragg SM, Beckham GT, Bruce NC et al (2015) Lignocellulose degradation mechanisms across the Tree of Life. Curr Opin Chem Biol 29:108–119. https://doi.org/10.1016/j.cbpa.2015.10.018

Currie MA, Adams JJ, Faucher F et al (2012) Scaffoldin conformation and dynamics revealed by a ternary complex from the Clostridium thermocellum cellulosome. J Biol Chem 287:26953–26961. https://doi.org/10.1074/jbc.M112.343897

Currie MA, Cameron K, Dias FMV et al (2013) Small angle X-ray scattering analysis of Clostridium thermocellum cellulosome N-terminal complexes reveals a highly dynamic structure. J Biol Chem 288:7978–7985. https://doi.org/10.1074/jbc.M112.408757

Dassa B, Borovok I, Lamed R et al (2012) Genome-wide analysis of acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. BMC Genom 13:210. https://doi.org/10.1186/1471-2164-13-210

Dassa B, Borovok I, Ruimy-Israeli V et al (2014) Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. PLoS ONE 9:e99221. https://doi.org/10.1371/journal.pone.0099221

Dassa B, Borovok I, Lombard V et al (2017) Pan-cellulosomics of mesophilic clostridia: variations on a theme. Microorganisms 5:74. https://doi.org/10.3390/microorganisms5040074

Davidi L, Moraïs S, Artzi L et al (2016) Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome. Proc Natl Acad Sci USA 113:10854–10859. https://doi.org/10.1073/pnas.1608012113

Demishtein A, Karpol A, Barak Y et al (2010) Characterization of a dockerin-based affinity tag: application for purification of a broad variety of target proteins. J Mol Recognit 23:525–535. https://doi.org/10.1002/jmr.1029

Doi RH, Kosugi A (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol 2:541–551. https://doi.org/10.1038/nrmicro925

Dror TW, Morag E, Rolider A et al (2003a) Regulation of the cellulosomal CelS (cel48A) gene of Clostridium thermocellum is growth rate dependent. J Bacteriol 185:3042–3048

Dror TW, Rolider A, Bayer EA et al (2003b) Regulation of expression of scaffoldin-related genes in Clostridium thermocellum. J Bacteriol 185:5109–5116

Dykstra AB, St Brice L, Rodriguez M et al (2014) Development of a multipoint quantitation method to simultaneously measure enzymatic and structural components of the Clostridium thermocellum cellulosome protein complex. J Proteome Res 13:692–701. https://doi.org/10.1021/pr400788e

Fan Z, Yuan L (2010) Production of multifunctional chimaeric enzymes in plants: a promising approach for degrading plant cell wall from within. Plant Biotechnol J 8:308–315. https://doi.org/10.1111/j.1467-7652.2009.00484.x

Fanutti C, Ponyi T, Black GW et al (1995) The Conserved Noncatalytic 40-Residue Sequence in Cellulases and Hemicellulases from Anaerobic Fungi Functions as a Protein Docking Domain. J Biol Chem 270:29314–29322. https://doi.org/10.1074/jbc.270.49.29314

Fierobe H-P, Bayer EA, Tardif C et al (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J Biol Chem 277:49621–49630. https://doi.org/10.1074/jbc.M207672200

Fierobe H-P, Mingardon F, Mechaly A et al (2005) Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J Biol Chem 280:16325–16334. https://doi.org/10.1074/jbc.M414449200

Fontes CMGA, Gilbert HJ (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem 79:655–681. https://doi.org/10.1146/annurev-biochem-091208-085603

Fujimoto Z, Jackson A, Michikawa M et al (2013) The structure of a *streptomyces avermitilis* α-l-rhamnosidase reveals a novel carbohydrate-binding module cbm67 within the six-domain arrangement. J Biol Chem 288:12376–12385. https://doi.org/10.1074/jbc.M113.460097

Fujino T, Béguin P, Aubert JP (1993) Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. J Bacteriol 175:1891–1899

Gefen G, Anbar M, Morag E et al (2012) Enhanced cellulose degradation by targeted integration of a cohesin-fused β-glucosidase into the Clostridium thermocellum cellulosome. Proc Natl Acad Sci USA 109:10298–10303. https://doi.org/10.1073/pnas.1202747109

Gerwig GJ, Kamerling JP, Vliegenthart JF et al (1993) The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of Clostridium thermocellum and Bacteroides cellulosolvens. J Biol Chem 268:26956–26960

Gilbert HJ, Knox JP, Boraston AB (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Curr Opin Struct Biol 23:669–677. https://doi.org/10.1016/j.sbi.2013.05.005

Gilkes NR, Warren RA, Miller RC, Kilburn DG (1988) Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis. J Biol Chem 263:10401–10407

Gilmore SP, Henske JK, O'Malley MA (2015) Driving biomass breakdown through engineered cellulosomes. Bioengineered 6:204–208. https://doi.org/10.1080/21655979.2015.1060379

Gilmore SP, Lillington SP, Haitjema CH et al (2020) Designing chimeric enzymes inspired by fungal cellulosomes. Synthetic and Systems Biotechnology 5:23–32. https://doi.org/10.1016/j.synbio.2020.01.003

Gold ND, Martin VJJ (2007) Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. J Bacteriol 189:6787–6795. https://doi.org/10.1128/JB.00882-07

Guillén D, Sánchez S, Rodríguez-Sanoja R (2009) Carbohydrate-binding domains: multiplicity of biological roles. Appl Microbiol Biotechnol 85:1241–1249. https://doi.org/10.1007/s00253-009-2331-y

Hagen LH, Brooke CG, Shaw C et al (2020) Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber. Microbiology

Haimovitz R, Barak Y, Morag E et al (2008) Cohesin-dockerin microarray: diverse specificities between two complementary families of interacting protein modules. Proteomics 8:968–979. https://doi.org/10.1002/pmic.200700486

Haitjema CH, Gilmore SP, Henske JK et al (2017) A parts list for fungal cellulosomes revealed by comparative genomics. Nature Microbiology 2:17087. https://doi.org/10.1038/nmicrobiol.2017.87

Hamberg Y, Ruimy-Israeli V, Dassa B et al (2014) Elaborate cellulosome architecture of Acetivibrio cellulolyticus revealed by selective screening of cohesin-dockerin interactions. PeerJ 2:e636. https://doi.org/10.7717/peerj.636

Hammel M, Fierobe H-P, Czjzek M et al (2004) Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. J Biol Chem 279:55985–55994. https://doi.org/10.1074/jbc.M408979200

Hammel M, Fierobe H-P, Czjzek M et al (2005) Structural basis of cellulosome efficiency explored by small angle X-ray scattering. J Biol Chem 280:38562–38568. https://doi.org/10.1074/jbc.M503168200

Han SO, Yukawa H, Inui M, Doi RH (2003) Regulation of expression of cellulosomal cellulase and hemicellulase genes in Clostridium cellulovorans. J Bacteriol 185:6067–6075

Hasunuma T, Okazaki F, Okai N et al (2013) A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol 135:513–522. https://doi.org/10.1016/j.biortech.2012.10.047

Henrissat B, Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 7:637–644. https://doi.org/10.1016/S0959-440X(97)80072-3

Henrissat B, Davies GJ (2000) Glycoside Hydrolases and Glycosyltransferases. Families, Modules, and Implications for Genomics. Plant Physiol 124:1515–1519. https://doi.org/10.1104/pp.124.4.1515

Himmel ME, Bayer EA (2009) Lignocellulose conversion to biofuels: current challenges, global perspectives. Curr Opin Biotechnol 20:316–317. https://doi.org/10.1016/j.copbio.2009.05.005

Himmel ME, Xu Q, Luo Y et al (2010) Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels 1:323–341. https://doi.org/10.4155/bfs.09.25

Hong W, Zhang J, Feng Y et al (2014) The contribution of cellulosomal scaffoldins to cellulose hydrolysis by Clostridium thermocellum analyzed by using thermotargetrons. Biotechnol Biofuels 7:80. https://doi.org/10.1186/1754-6834-7-80

Hyeon JE, Jeon WJ, Whang SY, Han SO (2011) Production of minicellulosomes for the enhanced hydrolysis of cellulosic substrates by recombinant corynebacterium glutamicum. Enzyme Microb Technol 48:371–377. https://doi.org/10.1016/j.enzmictec.2010.12.014

Hyeon JE, Kang DH, Han SO (2014) Signal amplification by a self-assembled biosensor system designed on the principle of dockerin-cohesin interactions in a cellulosome complex. Analyst 139:4790–4793. https://doi.org/10.1039/c4an00856a

Ichikawa S, Karita S, Kondo M, Goto M (2014) Cellulosomal carbohydrate-binding module from Clostridium josui binds to crystalline and non-crystalline cellulose, and soluble polysaccharides. FEBS Lett 588:3886–3890. https://doi.org/10.1016/j.febslet.2014.08.032

Israeli-Ruimy V, Bule P, Jindou S et al (2017) Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. Scientific Reports 7:42355. https://doi.org/10.1038/srep42355

Jørgensen H, Kristensen JB, Felby C (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioprod Biorefin 1:119–134. https://doi.org/10.1002/bbb.4

Kahel-Raifer H, Jindou S, Bahari L et al (2010) The unique set of putative membrane-associated anti-σ factors in Clostridium thermocellum suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation: biomass-sensing anti-σ factors in C. thermocellum. FEMS Microbiol Lett 308:84–93. https://doi.org/10.1111/j.1574-6968.2010.01997.x

Kahn A, Moraïs S, Galanopoulou AP, et al (2019) Creation of a functional hyperthermostable designer cellulosome. Biotechnol Biofuels 12:. https://doi.org/10.1186/s13068-019-1386-y

Kang S, Barak Y, Lamed R et al (2006) The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. Mol Microbiol 60:1344–1354. https://doi.org/10.1111/j.1365-2958.2006.05182.x

Karmakar M, Ray RR (2011) Current trends in research and application of microbial cellulases. Res J Microbiol 6:41–53. https://doi.org/10.3923/jm.2011.41.53

Karpol A, Kantorovich L, Demishtein A et al (2009) Engineering a reversible, high-affinity system for efficient protein purification based on the cohesin-dockerin interaction. J Mol Recognit 22:91–98. https://doi.org/10.1002/jmr.926

Kataeva IA, Uversky VN, Brewer JM et al (2004) Interactions between immunoglobulin-like and catalytic modules in Clostridium thermocellum cellulosomal cellobiohydrolase CbhA. Protein Eng Des Sel 17:759–769. https://doi.org/10.1093/protein/gzh094

Kim D-M, Nakazawa H, Umetsu M et al (2012) A nanocluster design for the construction of artificial cellulosomes. Catal Sci Technol 2:499. https://doi.org/10.1039/c2cy00371f

Kleine J, Liebl W (2006) Comparative characterization of deletion derivatives of the modular xylanase XynA of Thermotoga maritima. Extremophiles 10:373–381. https://doi.org/10.1007/s00792-006-0509-0

Kruus K, Lua AC, Demain AL, Wu JH (1995) The anchorage function of CipA (CelL), a scaffolding protein of the Clostridium thermocellum cellulosome. Proc Natl Acad Sci USA 92:9254–9258

Kuhad RC, Gupta R, Singh A (2011) Microbial cellulases and their industrial applications. Enzyme Res 2011:1–10. https://doi.org/10.4061/2011/280696

Kurokawa J, Hemjinda E, Arai T et al (2001) Sequence of the Clostridium thermocellum mannanase gene man26B and characterization of the translated product. Biosci Biotechnol Biochem 65:548–554

Lamed R, Setter E, Bayer EA (1983) Characterization of a cellulose-binding, cellulase-containing complex in Clostridium thermocellum. J Bacteriol 156:828–836

Leibovitz E, Béguin P (1996) A new type of cohesin domain that specifically binds the dockerin domain of the Clostridium thermocellum cellulosome-integrating protein CipA. J Bacteriol 178:3077–3084

Lemaire M, Ohayon H, Gounon P et al (1995) OlpB, a new outer layer protein of Clostridium thermocellum, and binding of its S-layer-like domains to components of the cell envelope. J Bacteriol 177:2451–2459

Levy-Assaraf M, Voronov-Goldman M, Rozman Grinberg I et al (2013) Crystal structure of an uncommon cellulosome-related protein module from Ruminococcus flavefaciens that resembles papain-like cysteine peptidases. PLoS ONE 8:e56138. https://doi.org/10.1371/journal.pone.0056138

Liang Y, Si T, Ang EL, Zhao H (2014) Engineered pentafunctional minicellulosome for simultaneous saccharification and ethanol fermentation in Saccharomyces cerevisiae. Appl Environ Microbiol 80:6677–6684. https://doi.org/10.1128/AEM.02070-14

Lombard V, Bernard T, Rancurel C et al (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J 432:437–444. https://doi.org/10.1042/BJ20101185

Lombard V, Golaconda Ramulu H, Drula E et al (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178

Mazmanian SK (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. Science 285:760–763. https://doi.org/10.1126/science.285.5428.760

Mesnage S (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface *via* a conserved mechanism involving wall polysaccharide pyruvylation. The EMBO J 19:4473–4484. https://doi.org/10.1093/emboj/19.17.4473

Mingardon F, Chanal A, Tardif C et al (2007) Exploration of new geometries in cellulosome-like chimeras. Appl Environ Microbiol 73:7138–7149. https://doi.org/10.1128/AEM.01306-07

Mitsuzawa S, Kagawa H, Li Y et al (2009) The rosettazyme: a synthetic cellulosome. J Biotechnol 143:139–144. https://doi.org/10.1016/j.jbiotec.2009.06.019

Molinier A-L, Nouailler M, Valette O et al (2011) Synergy, structure and conformational flexibility of hybrid cellulosomes displaying various inter-cohesins linkers. J Mol Biol 405:143–157. https://doi.org/10.1016/j.jmb.2010.10.013

Morag E, Halevy I, Bayer EA, Lamed R (1991) Isolation and properties of a major cellobiohydrolase from the cellulosome of Clostridium thermocellum. J Bacteriol 173:4155–4162

Morag E, Lapidot A, Govorko D et al (1995) Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of Clostridium thermocellum. Appl Environ Microbiol 61:1980–1986

Moraïs S, Morag E, Barak Y et al (2012) Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes. MBio 3. https://doi.org/10.1128/mBio.00508-12

Moraïs S, Shterzer N, Lamed R et al (2014) A combined cell-consortium approach for lignocellulose degradation by specialized lactobacillus plantarum cells. Biotechnol Biofuels 7:112. https://doi.org/10.1186/1754-6834-7-112

Moraïs S, Ben David Y, Bensoussan L et al (2016a) Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, ruminococcus champanellensis, reveals a fine-tuned system for cohesin-dockerin recognition. Environ Microbiol 18:542–556. https://doi.org/10.1111/1462-2920.13047

Moraïs S, Cockburn DW, Ben-David Y et al (2016b) Lysozyme activity of the ruminococcus champanellensis cellulosome. Environ Microbiol 18:5112–5122. https://doi.org/10.1111/1462-2920.13501

Mori Y, Ozasa S, Kitaoka M et al (2013) Aligning an endoglucanase Cel5A from thermobifida fusca on a DNA scaffold: potent design of an artificial cellulosome. Chem Commun (Camb) 49:6971–6973. https://doi.org/10.1039/c3cc42614a

Mosbah A, Belaïch A, Bornet O et al (2000) Solution structure of the module X2_1 of unknown function of the cellulosomal scaffolding protein CipC of Clostridium cellulolyticum. J Mol Biol 304:201–217. https://doi.org/10.1006/jmbi.2000.4192

Muñoz-Gutiérrez I, Ortiz de Ora L, Rozman Grinberg I et al (2016) Decoding biomass-sensing regulons of Clostridium thermocellum alternative sigma-I factors in a heterologous bacillus subtilis host system. PLoS ONE 11:e0146316. https://doi.org/10.1371/journal.pone.0146316

Nagy T, Tunnicliffe RB, Higgins LD et al (2007) Characterization of a double dockerin from the cellulosome of the anaerobic fungus Piromyces equi. J Mol Biol 373:612–622. https://doi.org/10.1016/j.jmb.2007.08.007

Nataf Y, Bahari L, Kahel-Raifer H et al (2010) Clostridium thermocellum cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proc Natl Acad Sci USA 107:18646–18651. https://doi.org/10.1073/pnas.1012175107

Ou J, Cao Y (2014) Incorporation of Nasutitermes takasagoensis endoglucanase into cell surface-displayed minicellulosomes in Pichia pastoris X33. J Microbiol Biotechnol 24:1178–1188

Pagès S, Bélaïch A, Bélaïch JP et al (1997) Species-specificity of the cohesin-dockerin interaction between Clostridium thermocellum and Clostridium cellulolyticum: prediction of specificity determinants of the dockerin domain. Proteins 29:517–527

Paget M (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules 5:1245–1265. https://doi.org/10.3390/biom5031245

Paterson GK, Mitchell TJ (2004) The biology of Gram-positive sortase enzymes. Trends in Microbiol 12:89–95. https://doi.org/10.1016/j.tim.2003.12.007

Peer A, Smith SP, Bayer EA et al (2009) Noncellulosomal cohesin—and dockerin-like modules in the three domains of life. FEMS Microbiol Lett 291:1–16. https://doi.org/10.1111/j.1574-6968.2008.01420.x

Perret S, Bélaich A, Fierobe H-P et al (2004) Towards designer cellulosomes in Clostridia: mannanase enrichment of the cellulosomes produced by Clostridium cellulolyticum. J Bacteriol 186:6544–6552. https://doi.org/10.1128/JB.186.19.6544-6552.2004

Pinheiro BA, Gilbert HJ, Sakka K et al (2009) Functional insights into the role of novel type I cohesin and dockerin domains from Clostridium thermocellum. Biochem J 424:375–384. https://doi.org/10.1042/BJ20091152

Pohlschröder M, Leschine SB, Canale-Parola E (1994) Multicomplex cellulase-xylanase system of Clostridium papyrosolvens C7. J Bacteriol 176:70–76. https://doi.org/10.1128/JB.176.1.70-76.1994

Poole D (1992) Identification of the cellulose-binding domain of the cellulosome subunit S1 from Clostridium thermocellum YS. FEMS Microbiol Lett 99:181–186. https://doi.org/10.1016/0378-1097(92)90022-G

Raghothama S, Eberhardt RY, Simpson P et al (2001) Characterization of a cellulosome dockerin domain from the anaerobic fungus piromyces equi. Nat Struct Biol 8:775–778. https://doi.org/10.1038/nsb0901-775

Raman B, Pan C, Hurst GB et al (2009) Impact of pretreated Switchgrass and biomass carbohydrates on Clostridium thermocellum ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS ONE 4:e5271. https://doi.org/10.1371/journal.pone.0005271

Ravachol J, Borne R, Tardif C et al (2014) Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium Clostridium cellulolyticum. J Biol Chem 289:7335–7348. https://doi.org/10.1074/jbc.M113.545046

Ravachol J, Borne R, Meynial-Salles I et al (2015) Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium ruminiclostridium cellulolyticum. Biotechnol Biofuels 8:114. https://doi.org/10.1186/s13068-015-0301-4

Ribeiro DO, Viegas A, Pires VMR et al (2019) Molecular basis for the preferential recognition of β1,3-1,4-glucans by the family 11 carbohydrate-binding module from Clostridium thermocellum. The FEBS J. https://doi.org/10.1111/febs.15162

Rincon MT, Martin JC, Aurilia V et al (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the ruminococcus flavefaciens 17 cellulosome. J Bacteriol 186:2576–2585. https://doi.org/10.1128/JB.186.9.2576-2585.2004

Rincon MT, Cepeljnik T, Martin JC et al (2005) Unconventional mode of attachment of the ruminococcus flavefaciens cellulosome to the cell surface. J Bacteriol 187:7569–7578. https://doi.org/10.1128/JB.187.22.7569-7578.2005

Rincon MT, Dassa B, Flint HJ et al (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, ruminococcus flavefaciens FD-1. PLoS ONE 5:e12476. https://doi.org/10.1371/journal.pone.0012476

Sabathé F, Bélaïch A, Soucaille P (2002) Characterization of the cellulolytic complex (cellulosome) of Clostridium acetobutylicum. FEMS Microbiol Lett 217:15–22

Salama-Alber O, Jobby MK, Chitayat S et al (2013) Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. J Biol Chem 288:16827–16838. https://doi.org/10.1074/jbc.M113.466672

Schneewind O, Model P, Fischetti VA (1992) Sorting of protein a to the staphylococcal cell wall. Cell 70:267–281. https://doi.org/10.1016/0092-8674(92)90101-H

Schubot FD, Kataeva IA, Chang J et al (2004) Structural basis for the exocellulase activity of the cellobiohydrolase cbha from Clostridium thermocellum †. Biochemistry 43:1163–1170. https://doi.org/10.1021/bi030202i

Shimon LJ, Pagès S, Belaich A et al (2000) Structure of a family IIIa scaffoldin CBD from the cellulosome of Clostridium cellulolyticum at 2.2 A resolution. Acta Crystallogr D Biol Crystallogr 56:1560–1568

Simpson PJ, Xie H, Bolam DN et al (2000) The Structural Basis for the Ligand Specificity of Family 2 Carbohydrate-binding Modules. J Biol Chem 275:41137–41142. https://doi.org/10.1074/jbc.M006948200

Sleytr UB, Messner P, Pum D, Sára M (1993) Crystalline bacterial cell surface layers. Mol Microbiol 10:911–916. https://doi.org/10.1111/j.1365-2958.1993.tb00962.x

Slutzki M, Reshef D, Barak Y et al (2015) Crucial roles of single residues in binding affinity, specificity, and promiscuity in the cellulosomal cohesin-dockerin interface. J Biol Chem 290:13654–13666. https://doi.org/10.1074/jbc.M115.651208

Smith SP, Bayer EA, Czjzek M (2017) Continually emerging mechanistic complexity of the multienzyme cellulosome complex. Curr Opin Struct Biol 44:151–160. https://doi.org/10.1016/j.sbi.2017.03.009

Stahl SW, Nash MA, Fried DB et al (2012) Single-molecule dissection of the high-affinity cohesin-dockerin complex. Proc Natl Acad Sci USA 109:20431–20436. https://doi.org/10.1073/pnas.1211929109

Steenbakkers PJ, Li XL, Ximenes EA et al (2001) Noncatalytic docking domains of cellulosomes of anaerobic fungi. J Bacteriol 183:5325–5333

Stern J, Kahn A, Vazana Y et al (2015) Significance of relative position of cellulases in designer cellulosomes for optimized cellulolysis. PLoS ONE 10:e0127326. https://doi.org/10.1371/journal.pone.0127326

Tamaru Y, Doi RH (2001) Pectate lyase A, an enzymatic subunit of the Clostridium cellulovorans cellulosome. Proc Natl Acad Sci USA 98:4125–4129. https://doi.org/10.1073/pnas.071045598

Tormo J, Lamed R, Chirino AJ et al (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J 15:5739–5751

Tsai S-L, Goyal G, Chen W (2010) Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. Appl Environ Microbiol 76:7514–7520. https://doi.org/10.1128/AEM.01777-10

Vazana Y, Barak Y, Unger T et al (2013) A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. Biotechnol Biofuels 6:182. https://doi.org/10.1186/1754-6834-6-182

Venditto I, Luis AS, Rydahl M et al (2016) Complexity of the Ruminococcus flavefaciens cellulosome reflects an expansion in glycan recognition. Proc Natl Acad Sci USA 113:7136–7141. https://doi.org/10.1073/pnas.1601558113

Voronov-Goldman M, Yaniv O, Gul O et al (2015) Standalone cohesin as a molecular shuttle in cellulosome assembly. FEBS Lett 589:1569–1576. https://doi.org/10.1016/j.febslet.2015.04.013

Wang Y, Leng L, Islam MK et al (2019) Substrate-related factors affecting cellulosome-induced hydrolysis for lignocellulose valorization. Int J Mol Sci 20:3354. https://doi.org/10.3390/ijms20133354

Willson BJ, Kovács K, Wilding-Steele T et al (2016) Production of a functional cell wall-anchored minicellulosome by recombinant Clostridium acetobutylicum ATCC 824. Biotechnol Biofuels 9:109. https://doi.org/10.1186/s13068-016-0526-x

Wilson Catriona A, Wood Thomas M (1992) The anaerobic fungus Neocallimastix frontalis: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose. Appl Microbiol Biotechnol 37. https://doi.org/10.1007/BF00174216

Xu Q, Gao W, Ding S-Y et al (2003) The cellulosome system of Acetivibrio cellulolyticus includes a novel type of adaptor protein and a cell surface anchoring protein. J Bacteriol 185:4548–4557

Xu Q, Barak Y, Kenig R et al (2004a) A novel Acetivibrio cellulolyticus anchoring scaffoldin that bears divergent cohesins. J Bacteriol 186:5782–5789. https://doi.org/10.1128/JB.186.17.5782-5789.2004

Xu Q, Bayer EA, Goldman M et al (2004b) Architecture of the Bacteroides cellulosolvens cellulosome: description of a cell surface-anchoring scaffoldin and a family 48 cellulase. J Bacteriol 186:968–977

Xu C, Huang R, Teng L et al (2015) Cellulosome stoichiometry in Clostridium cellulolyticum is regulated by selective RNA processing and stabilization. Nat Commun 6:6900. https://doi.org/10.1038/ncomms7900

Xu Q, Resch MG, Podkaminer K et al (2016) Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities. Sci Adv 2:e1501254. https://doi.org/10.1126/sciadv.1501254

You C, Zhang X-Z, Sathitsuksanoh N et al (2012) Enhanced microbial utilization of recalcitrant cellulose by an ex vivo cellulosome-microbe complex. Appl Environ Microbiol 78:1437–1444. https://doi.org/10.1128/AEM.07138-11

Ze X, Duncan SH, Louis P, Flint HJ (2012) Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. The ISME J 6:1535–1543. https://doi.org/10.1038/ismej.2012.4

Ze X, Ben David Y, Laverde-Gomez JA et al (2015) Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic firmicutes bacterium ruminococcus bromii. mBio 6. https://doi.org/10.1128/mBio.01058-15

Zhivin O, Dassa B, Moraïs S, et al (2017) Unique organization and unprecedented diversity of the Bacteroides (Pseudobacteroides) cellulosolvens cellulosome system. Biotechnology for Biofuels 10. https://doi.org/10.1186/s13068-017-0898-6

Chapter 10
Leucine Dehydrogenase: Structure and Thermostability

Hiroki Yamaguchi, Akiko Kamegawa, Kunio Nakata, Tatsuki Kashiwagi, Yoshinori Fujiyoshi, Kazutoshi Tani, and Toshimi Mizukoshi

Abstract Thermostability is a key factor in the industrial and clinical application of enzymes, and understanding mechanisms of thermostability is valuable for molecular biology and enzyme engineering. In this chapter, we focus on the thermostability of leucine dehydrogenase (LDH, EC 1.4.1.9), an amino acid-metabolizing enzyme that is an NAD^+-dependent oxidoreductase which catalyzes the deamination of branched-chain l-amino acids (BCAAs). LDH from *Geobacillus stearothermophilus* (*Gst*LDH) is a highly thermostable enzyme that has already been applied to quantify the concentration of BCAAs in biological specimens. However, the molecular mechanism of its thermostability had been unknown because no high-resolution structure was available. Here, we discuss the thermostability of *Gst*LDH on the basis of its structure determined by cryo-electron microscopy. Sequence comparison with other structurally characterized LDHs (from *Lysinibacillus sphaericus* and *Sporosarcina*

H. Yamaguchi · K. Nakata · T. Kashiwagi · T. Mizukoshi (✉)
AJINOMOTO Co., Inc, 1-1, Suzuki-Cho, Kawasaki-Ku, Kawasaki-Shi, Kanagawa 210-8681, Japan
e-mail: toshimi_mizukoshi@ajinomoto.com

H. Yamaguchi
e-mail: hiroki_yamaguchi@ajinomoto.com

K. Nakata
e-mail: kunio_nakata@ajinomoto.com

T. Kashiwagi
e-mail: tatsuki_kashiwagi@ajinomoto.com

A. Kamegawa · Y. Fujiyoshi
TMDU Advanced Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
e-mail: kame-ako.cesp@tmd.ac.jp

Y. Fujiyoshi
e-mail: yoshi.cesp@tmd.ac.jp

K. Tani
Graduate School of Medicine, Mie University, 2-174 Edobashi, Tsu, Mie 514-8507, Japan
e-mail: ktani@doc.medic.mie-u.ac.jp

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_10

psychrophila) indicated that non-conserved residues in *Gst*LDH, including Ala94, Tyr127, and the C-terminal region, are crucial for oligomeric stability through intermolecular interactions between protomers. Furthermore, NAD^+ binding to *Gst*LDH increased the thermostability of the enzyme as additional intermolecular interactions formed on cofactor binding. This knowledge is important for further applications and development of amino acid metabolizing enzymes in industrial and clinical fields.

Keywords Leucine dehydrogenase · Cryo-electron microscopy · NAD^+-dependent oxidoreductase · Thermostability · Amino acid metabolizing enzyme

Abbreviations

Cryo-EM	Cryo-electron microscopy
*Gst*LDH	LDH from *Geobacillus stearothermophilus*
FSC	Fourier shell correlation
LDH	Leucine dehydrogenase
*Lsp*LDH	LDH from *Lysinibacillus sphaericus*
NAD	Nicotinamide adenine dinucleotide
*Sps*LDH	LDH from *Sporosarcina psychrophila*
PDB	Protein Data Bank
PKU	Phenylketonuria
RMSD	Root-mean-square deviation

Introduction

Amino Acid Metabolizing Enzymes

Amino acids are vital compounds for living organisms and play a central role in metabolism (Wagner and Fell 2001). Most amino acids and related compounds are metabolized by a series of enzymes generally termed "amino acid metabolizing enzymes". The products are converted into various biomolecules in important pathways such as the tricarboxylic acid cycle, glutaminolysis, and the urea cycle. Thus, amino acid metabolizing enzymes play a crucial role in living systems. Dysfunction of amino acid metabolizing enzymes causes disorders such as hyperphosphatemia, hyperammonemia (MacMullen et al. 2001; Stanley et al. 2000, 1998), and phenylketonuria (PKU) (Huang et al. 1998; Nakamura et al. 1996; Rivero et al. 2000). Therefore, it is important to deepen our understanding of these amino acid metabolizing enzymes.

Furthermore, many amino acid metabolizing enzymes have been applied in industrial and clinical fields. In industrial applications, these enzymes make it possible

to manufacture various compounds with low stress on the environment because enzymes decrease the energy required for chemical reactions and catalyze them with high selectivity under mild conditions of temperature, pH, and pressure. Amino acid metabolizing enzymes have been used industrially in the manufacture of amino acids such as l-aspartic acid, l-cysteine and l-3,4-dihydroxyphenylalanine (l-DOPA) (Yokozeki 1988). In the clinical field, phenylalanine dehydrogenase can be used in colorimetric and electrochemical methods to quantify the concentration of phenylalanine (Huang et al. 1998; Nakamura et al. 1996), and this enzyme has been applied in mass screening of newborns for PKU (Nakamura et al. 1996; Tachibana et al. 2006; Naruse et al. 1992). The storage stability of enzymes is a key factor for their commercial and clinical development (Gibson 1999), and previous studies have reported that storage stability is correlated with enzyme thermostability (Cao et al. 2015). Therefore, it is meaningful to investigate the thermostability for further applications of amino acid metabolizing enzymes. In this chapter, we discuss the thermostability of leucine dehydrogenase (LDH; EC 1.4.1.9) on the basis of structural information, as an example of thermostabilization of an amino acid metabolizing enzyme.

Leucine Dehydrogenase

LDH belongs to the Glu/Leu/Phe/Val dehydrogenase family and reversibly catalyzes the oxidative deamination of l-leucine, l-isoleucine, and l-valine to their corresponding oxo-compounds using NAD^+ as a cofactor. Many studies have been conducted on LDH of *Bacillus* species. It plays a critical role in the metabolism of branched-chain l-amino acids (BCAAs) (Ohshima et al. 1978). This enzyme is applied in the quantification of BCAAs in a clinical context (Suzuki et al. 2008; Beckett et al. 1996). Patients with hepatic cirrhosis often exhibit a low concentration of BCAAs in the blood (Tajiri and Shimizu 2013). This induces a reduction of the molar ratio of BCAAs to tyrosine, and the Fisher ratio (the molar ratio of BCAAs to aromatic amino acids); therefore, BCAAs are widely used as an indicator of this disease (Campollo et al. 1992; Suzuki et al. 2008). In addition, LDH has also been used in stereoselective production of BCAAs in a bioreactor (Gu and Chang 1990a,1990b). The substrate specificity of LDH was changed by protein engineering for amine synthesis (Abrahamson et al. 2012). Thus, it is expected that LDH can be applied in a wide range of fields. For the sake of these applications, it is important to understand the structure–function relationships of LDH.

A number of studies have reported characterization of LDHs, particularly the highly thermostable type from *Geobacillus stearothermophilus* (*Gst*LDH) (Kataoka and Tanizawa 2003; Nagata et al. 1988; Matsuyama et al. 1992; Ohshima et al. 1985; Sekimoto et al. 1994), which is an essential enzyme for applications in clinical and industrial contexts. The LDHs from *Lysinibacillus sphaericus* (*Lsp*LDH; Protein Data Bank [PDB] ID: 1LEH) (Baker et al. 1995) and *Sporosarcina psychrophila* (*Sps*LDH; PDB ID: 3VPX) (Zhao et al. 2012) have been structurally characterized. Interestingly, the sequence similarity of *Gst*LDH is 96% with *Lsp*LDH and

94% with *Sps*LDH (Fig. 10.1), but the thermostability of *Gst*LDH, *Lsp*LDH, and *Sps*LDH are 70, 55, and 45 °C, respectively (Ohshima et al. 1978, 1985; Zhao et al. 2012). However, the mechanism of their thermostability was unknown until recently, because the structure of *Gst*LDH had not been elucidated. We have now structurally characterized *Gst*LDH by cryo-electron microscopy (cryo-EM), and in this chapter, we discuss the enhanced thermostability based on the structure of *Gst*LDH in comparison with those of the other LDHs.

Furthermore, there has been little research on the NAD^+-bound structure(s) of LDHs (*i.e.*, the holo form of the enzyme). The structures of *Lsp*LDH and *Sps*LDH were determined without cofactor NAD^+ (*i.e.* in the apo form). A previous study partially reported the NAD^+-bound *Lsp*LDH structure, but coordinates were not

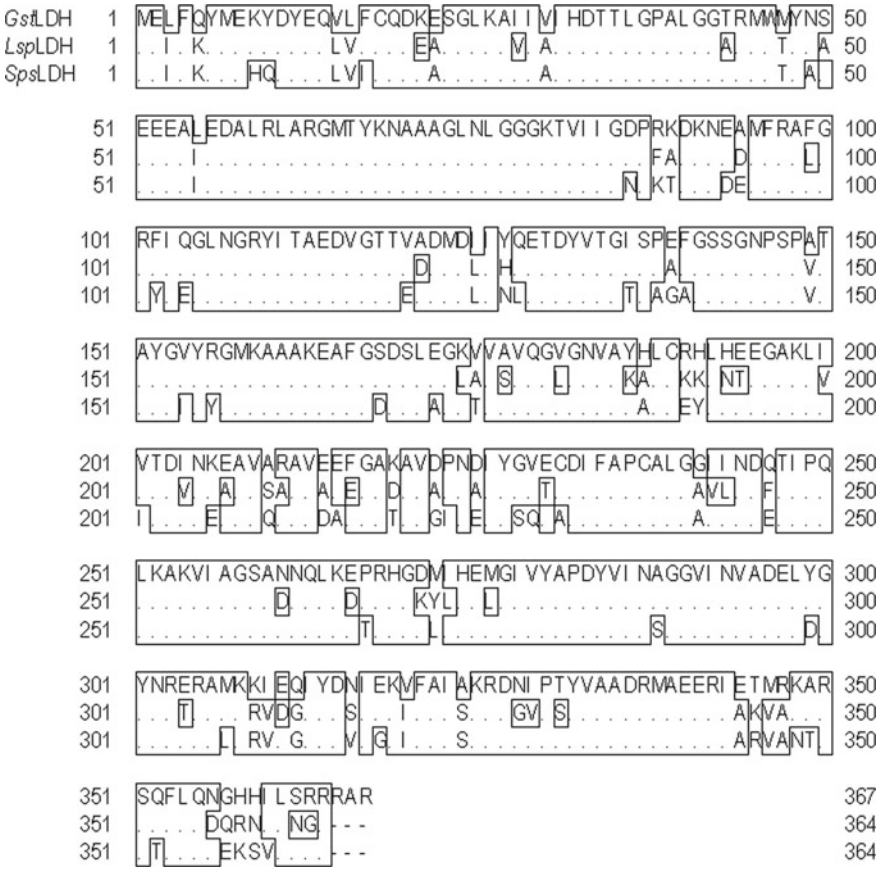


Fig. 10.1 Multiple sequence alignment of three leucine dehydrogenases. Amino acid sequence alignment of three structurally characterized leucine dehydrogenases (LDHs). The sequence of *Geobacillus stearothermophilus* LDH (*Gst*LDH) was used as the query and compared with the LDHs from *Lysinibacillus sphaericus* (*Lsp*LDH) and *Sporosarcina psychrophila* (*Sps*LDH). Dots indicate conserved residues and are boxed

deposited in the PDB (Baker et al. 1995). Here, we assess the NAD^+-bound structure of *Gst*LDH using cryo-EM data and computational analysis.

Structure of LDH from *G. Stearothermophilus*

We purified both apo and NAD^+-bound forms of recombinant *Gst*LDH based on our method (Yamaguchi et al. 2019). Cryo-EM data were collected using a 300-kV electron microscope, JEM-3000SFF with a helium stage (Fujiyoshi et al. 1991; Fujiyoshi 1998). Each image had an appropriate contrast, and the Thon rings were visible at ~3 Å resolution. Ring-shaped features were clearly observed in two-dimensional class averages. Three-dimensional maps of both apo (PDB ID: 6ACF and EMDataBank ID: EMD-9590) and NAD^+-bound forms (PDB ID: 6ACH and EMDataBank ID: EMD-9592) of *Gst*LDH were reconstructed from 132,800 and 92,034 particles, respectively, with D4 symmetry (Fig. 10.2 and Table 10.1). The final resolutions after refinement were 3.0 and 3.2 Å using the Fourier shell correlation (FSC) = 0.143 criterion (Rosenthal and Henderson 2003) (Fig. 10.2c). Previous studies have reported that *Gst*LDH was a hexamer, but our structure of *Gst*LDH was an octamer, which is consistent with other LDH structures (i.e., those of *Lsp*LDH and *Sps*LDH) (Zhao et al. 2012; Ohshima et al. 1985; Baker et al. 1995). Size exclusion chromatography analysis also supported an octameric form (Yamaguchi et al. 2019).

To clarify any structural differences among protomers of *Gst*LDH, focused classification using mask (Bai et al. 2015) and localized reconstruction with C1 symmetry were applied to the analyses of both the apo and NAD^+-bound forms using RELION (Scheres 2012; Bai et al. 2015; Roh et al. 2017). The atomic structures of both the apo and NAD^+-bound forms were also determined. The resolution of these maps was estimated to be 3.2 Å and 3.3 Å based on FSC = 0.143 (Rosenthal and Henderson 2003), respectively. The root-mean-square deviations (RMSD) of Cα atoms in the main chain between the structures from localized (C1 symmetry) and global (D4

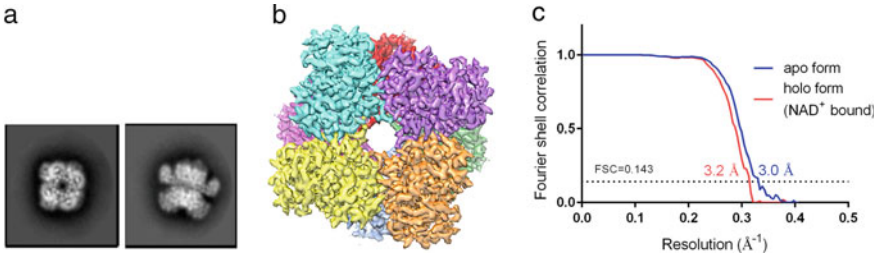


Fig. 10.2 Overall structure of *G. stearothermophilus* LDH based on cryo-electron microscopic analysis. **a** Two-dimensional class average. **b** Three-dimensional density map. Each protomer in the octamer of the apo form is shown in a different color. **c** Fourier shell correlation curve of apo and NAD^+-bound forms. The final resolutions were 3.0 Å for the apo (blue) and 3.2 Å for the NAD^+-complexed (red) forms

Table 10.1 Cryo-EM data collection, refinement, and validation statistics

	LDH (apo)	LDH (NAD^+)
Data collection		
Magnification	40,600	40,600
Voltage (kV)	300	300
Electron exposure (e-/Å^2)	42.6	42.6
Defocus range (μm)	−1.3 ~ −2.7	−1.2 ~ −2.8
Pixel size (Å) *	0.616	0.616
Symmetry imposed	D4	D4
Initial particle images (no.)	225,690	315,024
Final particle images (no.)	132,800	92,034
Map resolution (Å)	3.03	3.20
FSC threshold	0.143	0.143
Refinement and validation		
Model resolution (Å)	3.03	3.20
FSC threshold	0.143	0.143
Model resolution range (Å)	236.54–3.0	236.54–3.2
Map sharping *B*-factor (Å^2)	−122.8	−124.8
No. of atoms		
Protein	23,144	22,888
Ligand	0	352
B-factors (Å^2)		
Protein	44.0	49.1
Ligand	N/A	7.9
R.m.s deviations		
Bond lengths (Å)	0.008	0.013
Bond angles (°)	0.935	1.148
Ramachandran plot		
Favored (%)	95.08	92.78
Allowed (%)	4.93	7.22
Disallowed (%)	0.0	0.0

*Pixel size for the super-resolution mode

symmetry) reconstructions were 0.160 Å and 0.171 Å for the apo and holo forms, respectively. Thus, in the following text, our discussion is based on the structures with D4 symmetry.

In the structural comparison, extensive conformational change was not observed between the apo and NAD^+-bound forms. The RMSD of Cα atoms in the main chain was 0.182 Å and 0.157 Å for the octamer and protomer, respectively. *Gst*LDH had no allosteric regulation on NAD^+ binding; the Hill coefficient (Hill 1910) for NAD^+ was 0.95. Domain I (residues 1–136 and 332–367) is involved in the octamer formation

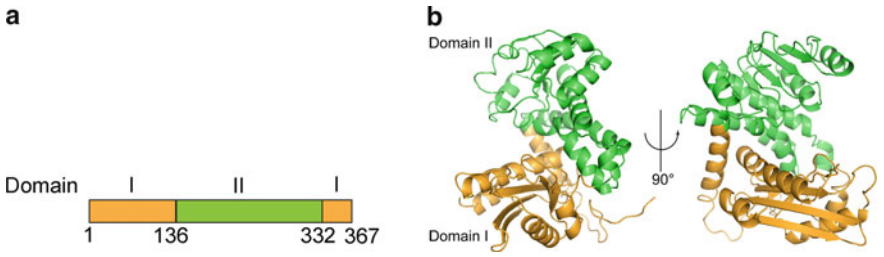


Fig. 10.3 Protomer of *G. stearothermophilus* LDH. **a** Schematic representation of the domains of LDH from *G. stearothermophilus*. **b** Domain I (orange) and domain II (green) in the protomer structure of the apo form of *Gst*LDH are shown in a cartoon model

of *Gst*LDH (Fig. 10.3). NAD^+ bound to domain II (residues 137–331), which is the nucleotide binding domain with a six stranded parallel β-sheet surrounded by four α-helices, including a dinucleotide-binding motif. On comparison between each domain in the apo and NAD^+-bound forms, the B-factors of the nucleotide binding domain (domain II) were higher than those of domain I (Fig. 10.4a). NAD^+ binding increased the B-factors of domain II without large conformational change (Fig. 10.4). However, the B-factors of domain I in the NAD^+-bound form were slightly lower than those in the apo form. The NAD^+-bound form had a disordered region (residues 142–144) in the loop between domains I and II (Fig. 10.4a). The resolution of the NAD^+ bound form might be decreased (Fig. 10.2c) by the structural changes that occur on NAD^+ binding. NAD^+ binding makes domain I more rigid but domain II more flexible.

Thermostability of *Gst*LDH

*Gst*LDH has the highest thermostability among the structurally characterized LDHs (Zhao et al. 2012; Baker et al. 1995). As mentioned above, the thermostability of *Gst*LDH, *Lsp*LDH, and *Sps*LDH were 70, 55, and 45 °C, respectively (Ohshima et al. 1978, 1985; Zhao et al. 2012). Interestingly, the sequence similarity of *Gst*LDH was 96% with *Lsp*LDH and 94% with *Sps*LDH (Fig. 10.1). This indicated that minor differences in amino acid sequence contribute to quite different thermostabilities. We evaluated 20 residues that are different among the three LDHs (Fig. 10.1). Residues of *Gst*LDH were mutated to the corresponding residues in *Sps*LDH and the effect on thermostability was determined. Notably, mutations Ala94Glu and Tyr127Asn decreased the thermostability of *Gst*LDH. The residual activity after treatment at 70 °C for 10 min was 19 and 29% for the Ala94Glu and Tyr127Asn variants, respectively (Yamaguchi et al. 2019). Thus, we assessed stabilization mechanisms of residues Ala94 and Tyr127 of *Gst*LDH from a structural point of view.

First, we analyzed the interactions made by Ala94 in *Gst*LDH. Ala94 undergoes a hydrophobic interaction with Ala349 in a neighboring protomer (Fig. 10.5a).

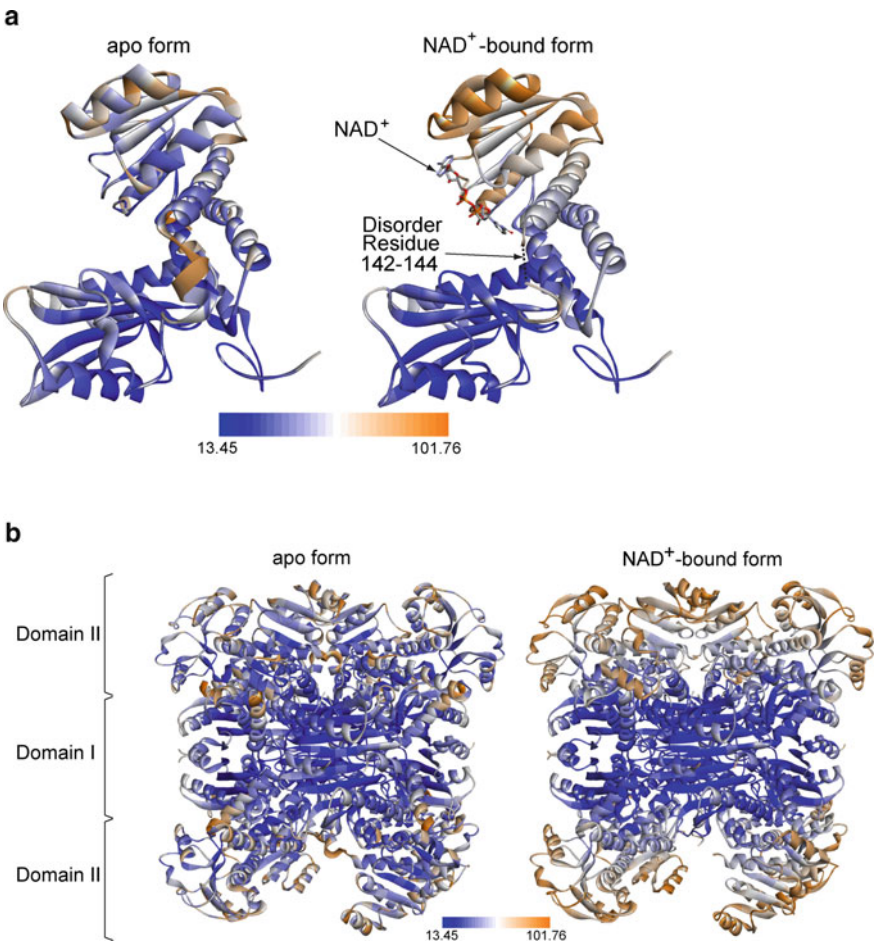


Fig. 10.4 B-factors in LDH of *G. stearothermophilus*. B-factors in the structural model of the apo (left) and NAD^+-bound (right) forms. The cartoon models are colored from blue to orange according to the B-factor values. **a** The protomer and **b** the overall structures

Ala94 and Ala349 are buried by surrounding residues, which form a hydrophobic patch (Fig. 10.5c). It is well-known that various proteins from thermophiles use hydrophobic interactions at their interfaces between oligomers to acquire thermostability (Vieille and Zeikus 2001). Furthermore, a number of studies have shown that thermostability was improved by exchanging hydrophilic residues for hydrophobic residues in the interface or inner regions (Haney et al. 1999; Kirino et al. 1994). Ala94 and Ala349 are located in the hydrophobic patch of the oligomer interface of *Gst*LDH and may have an important role in oligomerization. Our findings indicate that hydrophobic packing between protomers via domain I contributes to the high thermostability of *Gst*LDH. However, in *Sps*LDH, Glu94 forms an electrostatic

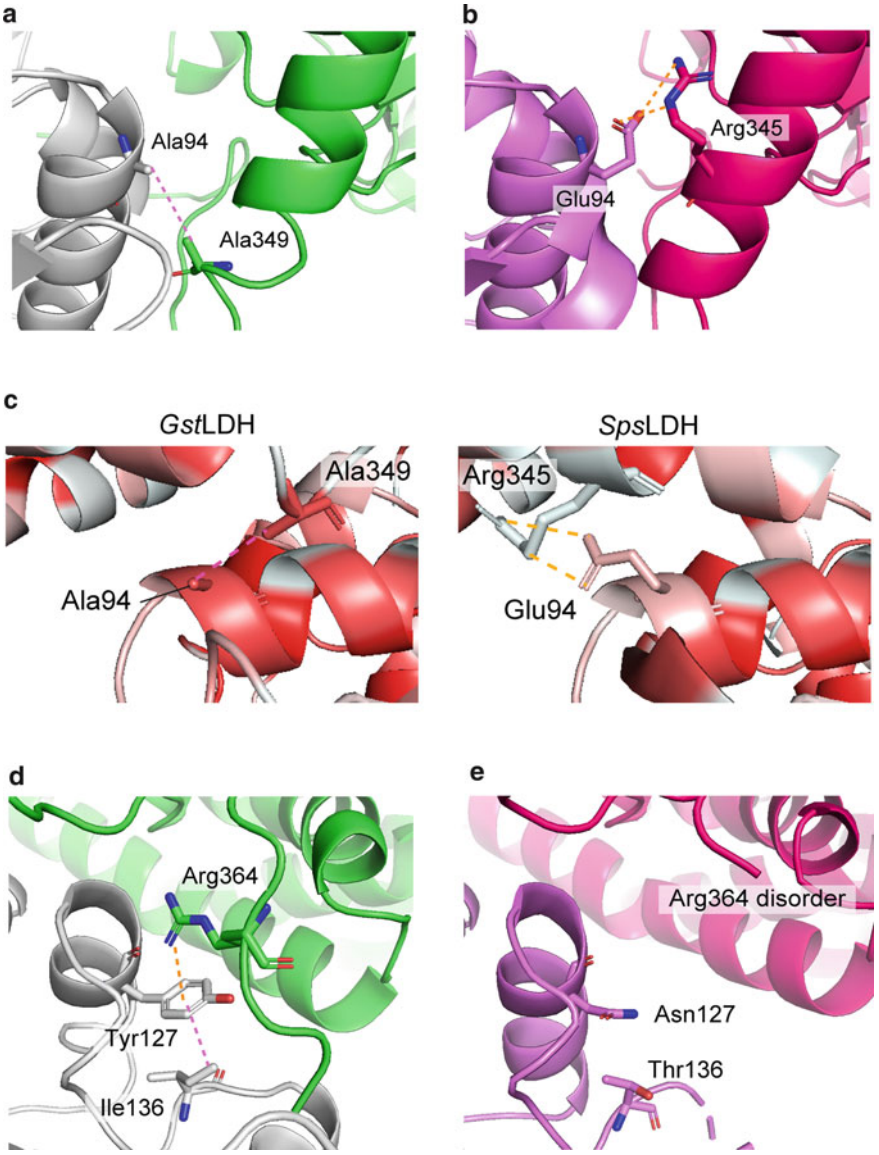


Fig. 10.5 Interactions between adjacent chains of LDHs of *G. stearothermophilus* (*Gst*LDH) and *S. psychrophila* (*Sps*LDH). Chains A (white) and G (green) in the apo form of *Gst*LDH and chains A (purple) and G (pink) in *Sps*LDH are shown as ribbon models. Interactions are shown by dotted lines (pink, hydrophobic interactions; orange, electrostatic interactions). Each interaction was detected using Discovery Studio 2017 R2 (BIOVIA) software. **a** The side chain of Ala94 in *Gst*LDH interacts with Ala349 in a neighboring protomer. **b** The side chain of Ala94 in *Sps*LDH interacts with Arg345. **c** Hydrophobicity of the residues around Ala94 (*Gst*LDH, left) and Glu94 (*Sps*LDH, right). The cartoon models are colored from white to red according to the hydrophobicity. **d** The side chain of Tyr127 in *Gst*LDH interacts with Ile136 in the same protomer, and Arg364 in a neighboring protomer. **e** The side chain of Asn127 in *Sps*LDH makes no interactions

interaction with Arg345, and these residues disrupt the hydrophobic patch (Fig. 10.5b, c).

In *Gst*LDH, the side chain of Tyr127 undergoes hydrophobic interaction with Ile136 in the same subunit, and an electrostatic interaction with Arg364 in a neighboring protomer (Fig. 10.5d), while Asn127 in *Sps*LDH makes no interaction with other residues (Fig. 10.5e). *Gst*LDH thus has additional interactions between protomers compared with *Sps*LDH. Our data therefore indicate that residues Ala94 and Tyr127 contribute to the thermostability of *Gst*LDH by improving the hydrophobicity and intermolecular interactions at the oligomeric interface.

A notable difference among the primary structures of the three LDHs was found in the C-terminal region; *Gst*LDH has a three amino acid (Arg365–Ala366–Arg367) longer tail than *Lsp*LDH and *Sps*LDH (Fig. 10.1 and Fig. 10.6a). In addition, the position of Arg364 was quite different among the three LDHs (Fig. 10.6a). Interestingly, Arg364 of *Sps*LDH was disordered in the structure (Zhao et al. 2012). To investigate the role of Arg364 and the C-terminus, alanine mutants (Arg364Ala, Arg365Ala, and Arg367Ala) of *Gst*LDH were characterized using the same method as mentioned above. The three alanine mutants showed lower residual activity than the wild-type (Fig. 10.6b). Interestingly, Arg364Ala had only 5% residual activity, lower than that of Arg365Ala and Arg367Ala (Fig. 10.6b). The side chain of Arg364 interacts with that of Tyr127, which is a key residue for the thermostability of *Gst*LDH though its intermolecular interaction with Ile136 (see above) (Fig. 10.5c). Therefore, these data indicated that the interactions among three residues (Tyr127, Ile136, and Arg364) are essential for the thermostability of *Gst*LDH. The fact that the residual activity of an Ile136Ala mutant also decreased (Fig. 10.6b) supported this speculation. Arg365 and Arg367 interact with each other without disordering the terminus structure, and then contribute to fixing the position of Arg364 to interact with Tyr127.

NAD^+ Recognition by Gst LDH

Structural Comparisons with Other LDHs

The structures of *Gst*LDH were determined by cryo-EM of apo and NAD^+-bound forms. However, the available structures of *Lsp*LDH and *Sps*LDH are only in the apo form, determined by X-ray crystallography (Zhao et al. 2012; Baker et al. 1995). A previous report discussed the NAD^+-bound structure of *Lsp*LDH (Baker et al. 1995), but the structure was not registered in the PDB. Here, the difference between *Gst*LDH and *Lsp*LDH is examined.

First, the apo forms of *Gst*LDH and *Lsp*LDH were compared. Two subunits (chain A: open form, and B: closed form) in the asymmetric unit of *Lsp*LDH had slightly different conformations; the RMSD of Cα atoms between chains A and B of *Lsp*LDH was 0.700 Å. However, each protomer of *Gst*LDH had the same conformation. To clarify the conformational differences, the movement of domains I and II was

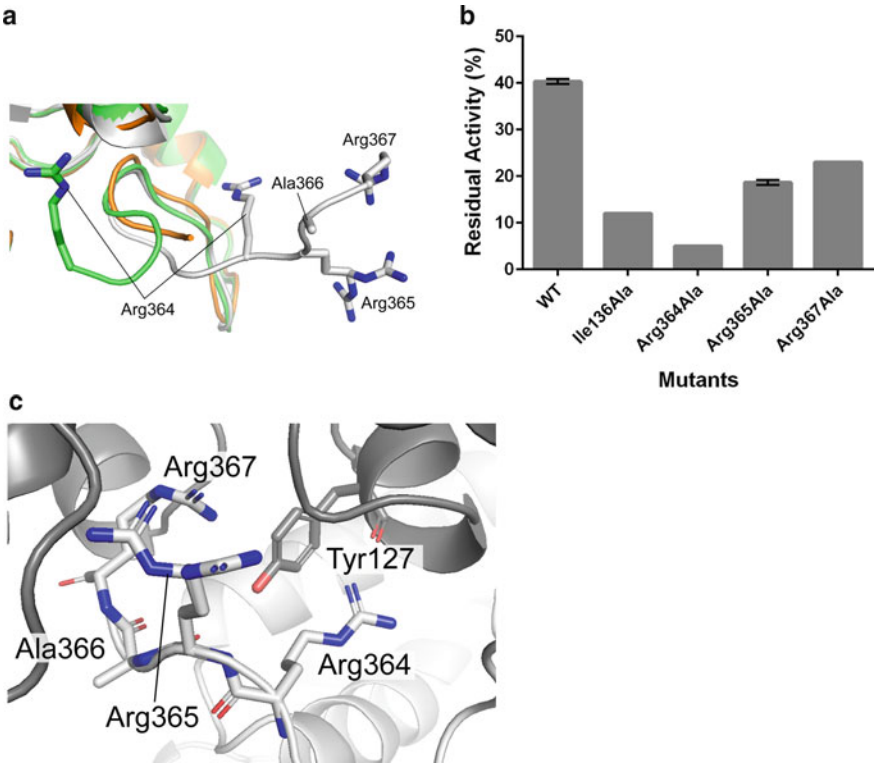


Fig. 10.6 The C-terminal region of three LDHs. **a** The C-terminal region of *G. stearothermophilus* LDH (*Gst*LDH) is shown in white as a ribbon model, including residues Arg364, Arg365, Ala366 and Arg367. The structures of *L. sphaericus* LDH (green), and *Sporosarcina psychrophila* (orange) contain Arg364, but lack the other residues. Arg364 of *Sps*LDH disappears due to disorder. **b** Relative activities of wild-type and site-directed mutants of *Gst*LDH after incubation at 70 °C for 10 min. The residual activities were calculated with each activity before incubation defined as 100% (N = 3, mean ± S.D.). **c** Density map of Tyr127 and the C-terminus of *Gst*LDH. Chain A (white) and a neighboring chain (C, in gray) in *Gst*LDH are shown as a cartoon model. The side chains of Tyr127, Arg364, Arg365, Ala366, and Arg367 are shown as stick models

analyzed along the hinge axis defined by the Cα atoms of residues 68, 72, 134, 297, 329, and 335 mentioned in a previous study of *Lsp*LDH after each domain II was superimposed (Baker et al. 1995). The conformation of *Gst*LDH determined by cryo-EM was close to that of chain A in *Lsp*LDH determined by X-ray crystallography (Fig. 10.7); the differences in domain hinge angle were <1.3°. However, the difference of the hinge angle between *Gst*LDH and chain B of *Lsp*LDH was 3.2°, with domain II bending toward domain I to form a smaller cleft for the catalytic site.

Next, on comparison of the apo and NAD^+-bound forms of *Gst*LDH, the RMSD was 0.157 Å, less than the RMSD for the comparison between *Gst*LDH and the chain A (0.614 Å) and chain B (0.636 Å) of *Lsp*LDH. The NAD^+-bound form of *Gst*LDH showed that one molecule of NAD^+ is bound to each protomer and the

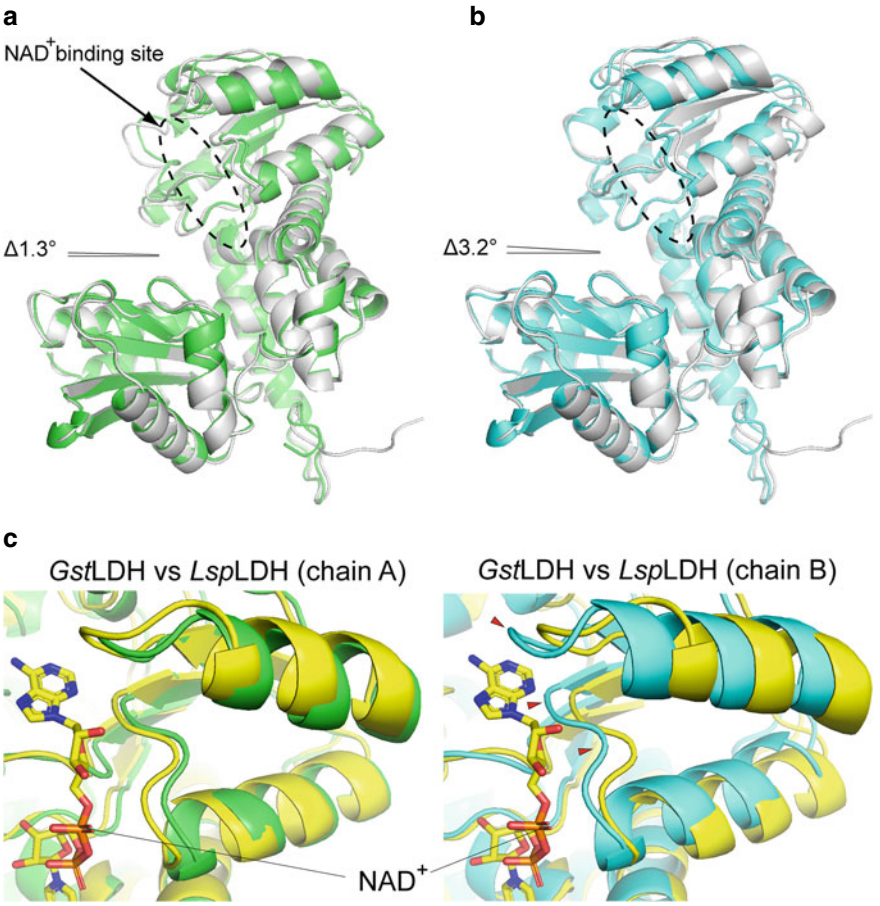


Fig. 10.7 Structural comparison between the LDHs from *G. stearothermophilus* (*Gst*LDH) and *L. sphaericus (Lsp*LDH*).* (**a**, **b**) The structures of apo *Gst*LDH (white*)* and apo *Lsp*LDH (chain A in green, chain B in cyan) superimposed based on Cα atoms in the main chain. The movement can be analyzed relative to the hinge axis defined by the Cα atoms of residues 68, 72, 134, 297, 329, and 335 of *Lsp*LDH (shown in a previous study(Baker et al. 1995)). The difference in the domain angle between *Gst*LDH and chain A (**a**) and chain B (**b**) of *Lsp*LDH was 1.3° and 3.2°, respectively. The dotted circle is the NAD^+ binding site. (**c**) The NAD^+-bound form of *Gst*LDH (yellow) and apo form of *Lsp*LDH (chain A in green, chain B in cyan) superimposed based on Cα atoms in the main chain. The NAD^+ binding region is shown. NAD^+ is shown as a stick model. The region of structural difference is shown as red triangles

density of NAD^+ was clearly observed in the active site of domain II (Fig. 10.8). There was no difference between the eight protomers. Thus, the conformation of NAD^+-bound *Gst*LDH was similar to both the apo form of *Gst*LDH and of chain A in *Lsp*LDH. This fact corresponded well with the previous report that NAD^+ density was observed only in chain A (the "open" form) of *Lsp*LDH, but no cofactor was

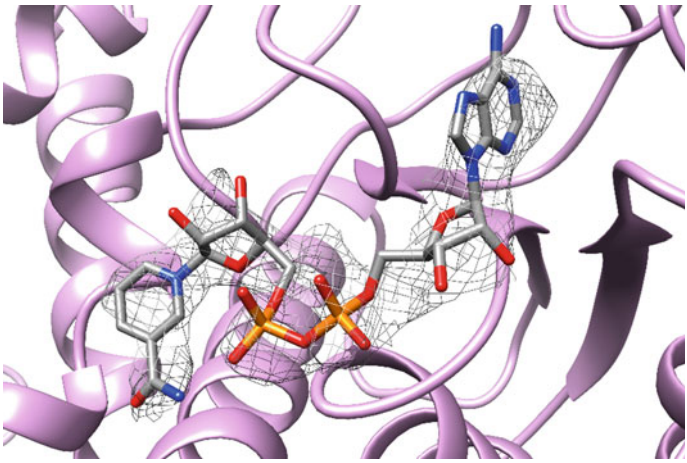


Fig. 10.8 Density map of NAD^+ in the NAD^+-bound form of *G. stearothermophilus* LDH. The density shown as mesh (gray) corresponds to one NAD^+ molecule. Protein and NAD^+ are drawn as a cartoon (pink) and a stick model (white), respectively. The density map within 1.96 Å is shown at level 0.04 using UCSF Chimera

observed in chain B (the "closed" form) (Baker et al. 1995). It was also reported that soaking of *Lsp*LDH crystals in NAD^+ affected the crystalline structure, and that a high concentration of NAD^+ reduced the quality of the diffraction data. A movement of domain II of chain B of *Lsp*LDH in the crystal structure was probably caused by crystal packing, and is likely to interfere with NAD^+ binding (Fig. 10.7c) during soaking of NAD^+ into crystals.

NAD^+ Interaction with GstLDH

Interactions between NAD^+ and the protomer of *Gst*LDH were analyzed computationally, using Discovery Studio 2017 R2 (BIOVIA) software (Fig. 10.9). The methyl groups of Ile204 and Ala238 recognized the adenine ring of NAD^+ by pi–alkyl interaction. The main chain of Gly180 and the side chain of Asp203 form carbonyl hydrogen bonds to the ribose moiety. In addition, the ribose forms hydrogen bonds with the side chain of Ser259 and the main chains of both Cys237 and Ala238. The amino group of Asn261 is located near the 2′-hydroxyl group of ribose and forms a Van der Waals interaction. The main chain of Ser259 interacts with the nicotinamide ring via a hydrogen bond. The carboxamide group forms a hydrogen bond with the hydroxyl groups of Ser147 and Thr150. An oxygen atom in the phosphate moiety forms a hydrogen bond to the main chain of Gly182.

By single particle analysis, NAD^+ bound all *Gst*LDH protomers in the same manner. However, a previous report showed that only one or two molecules in the

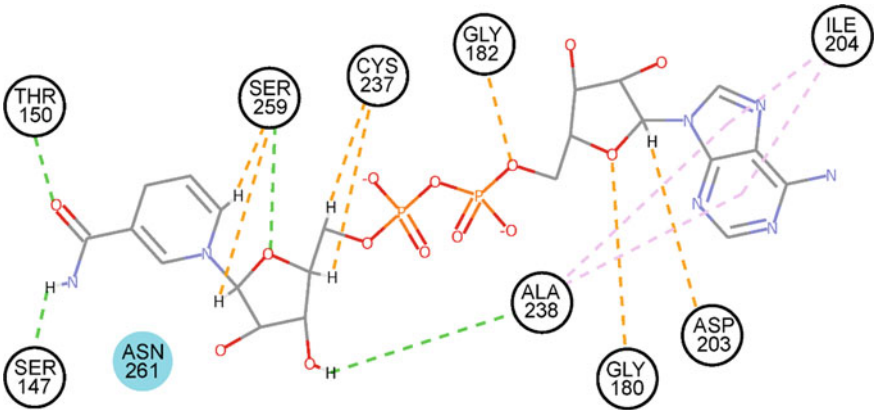


Fig. 10.9 Interaction between NAD^+ and *G. stearothermophilus* LDH (*Gst*LDH). Two-dimensional diagram of the interactions between NAD^+ and *Gst*LDH, analyzed using Discovery Studio 2017 R2 (BIOVIA). Each interaction is shown as a dotted line (orange, hydrogen bond; green, conventional hydrogen bond; blue, Van der Waals interaction; pink, pi–alkyl interaction)

asymmetric unit of *Lsp*LDH interacted with NAD^+ (Baker et al. 1995). The residues close to the NAD^+ binding site were conserved in both structures. Eight residues (Ser147, Thr150, Gly180, Ile204, Ala238, Leu239, Ser259, and Asn261) also interacted with NAD^+ in *Lsp*LDH. The interactions of Gly182 and Cys237 were only observed in cofactor binding by *Gst*LDH. Gln179, Asn183, and Pro223 interact with NAD^+ in the NAD^+-bound structure of *Lsp*LDH only. The differences in NAD^+ recognition between *Gst*LDH and *Lsp*LDH might be attributed to the analyzing methods used, *i.e.* X-ray crystallography or cryo-EM.

Thermostabilization by NAD^+ Binding

Based on the comparison of B-factors between the apo and NAD^+-bound forms of *Gst*LDH (Fig. 10.4), it was concluded that NAD^+ binding stabilizes the structure of domain I, which is related to octamer formation. In fact, the residual activity of the NAD^+-bound form after incubation at 80 °C was three times that of the apo form (Yamaguchi et al. 2019). We considered the relationship between thermostabilization and octamer formation by comparison of the apo and NAD^+-bound structures. The nicotinamide moiety of NAD^+ interacted with Ser147 through a hydrogen bond (Fig. 10.9). Ser147 approached NAD^+ by 0.5 Å and Pro146 also moved 1.0 Å toward the cofactor (Fig. 10.10a) relative to their positions in the apo structure. These changes pull Asn145 in the same direction, and induce a 1.9 Å movement of its Cα atom. On NAD^+ binding, conformational change of residues 145–147 triggers disorder of the neighboring residues 142–144 (Fig. 10.10a). The disorder of residues 142–144 perturbed the Cα of residues 136–141. As a result, Ile136 moved toward NAD^+ by

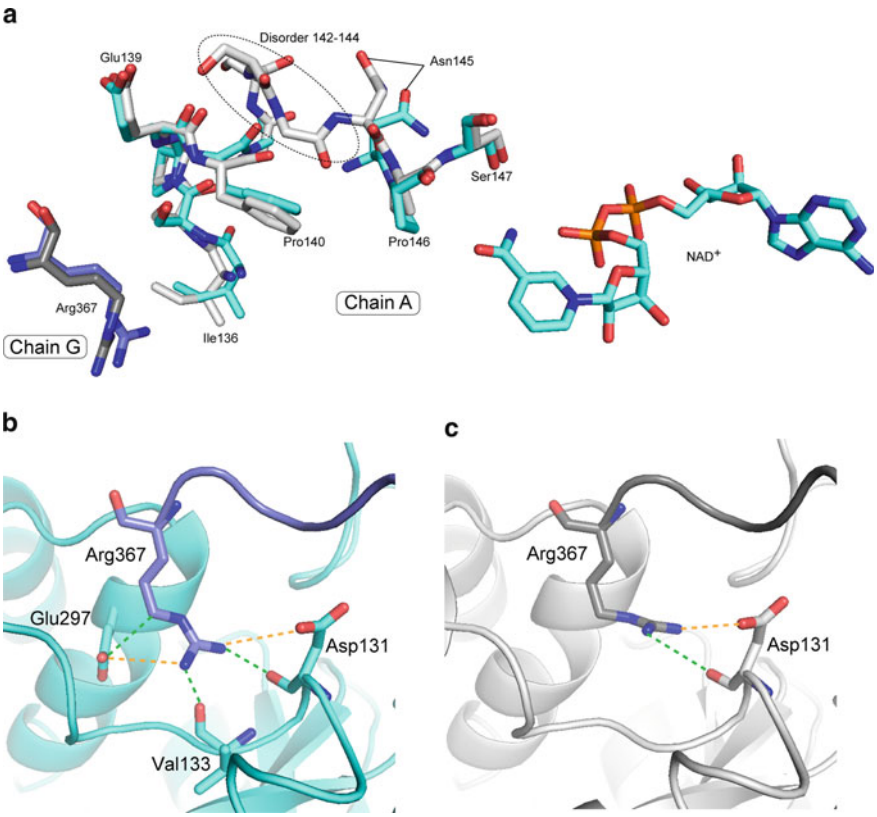


Fig. 10.10 Conformational change occurring in *G. stearothermophilus* LDH on NAD^+ binding. **a** Superimposed structures of the apo and NAD^+-bound forms: chain A (white and light green, respectively), and neighboring chain G (gray and blue, respectively). The dotted circle shows the disordered residues (142–144) in the NAD^+-bound form. (**b**, **c**) The interactions between chain A and Arg367 of chain G were analyzed using Discovery Studio 2017 R2 (BIOVIA) software. The interactions are shown as dotted lines (green, hydrogen bond; orange, electrostatic interactions). Each interacting residue is drawn as a stick model. **b** In the NAD^+-bound form, the side chain of Arg367 interacts with Asp131, Val133, and Glu297; **c** in the apo form, the side chain of Arg367 interacts with Asp131

1.1 Å and this motion extended the space between Ile136 of chain A and Arg367 of the neighboring protomer, chain G. The side chain of Arg367 changed conformation and shifted in the same direction (i.e. toward the new position of Ile136). Consequently, Arg367 acquired an electrostatic interaction with the side chain of Glu297 and hydrogen bonds to the carbonyl oxygen of Val133 in domain I (Fig. 10.10b), while in the apo form, Arg367 had no interactions with Glu297 or Val133 (Fig. 10.10c). Therefore, the interaction between NAD^+ and Ser147 transduced a series of conformational changes that increased intermolecular interactions in the oligomer interface.

This mechanism may explain the increase in thermostability of *Gst*LDH on NAD^+ binding.

Conclusions

In this chapter, we have discussed the thermostability of LDH on the basis of structural data for *Gst*LDH determined by cryo-EM. The results of comparison with other LDHs show that residues Ala94, Tyr127 and the C-terminal region (Arg364 to Arg367) play an important role in determining the high thermostability of *Gst*LDH through intermolecular interactions between protomers. In addition, it was found that NAD^+ binding to domain II of *Gst*LDH increased the thermostability of the enzyme by creating extra intermolecular interactions in domain I. *Gst*LDH has high thermostability because of stabilization of the oligomer via interactions at the oligomer interface. This knowledge is valuable for further applications of the enzyme, and others in the Glu/Leu/Phe/Val dehydrogenase family, in industrial and clinical fields, and will help in improvement of thermostability by enzyme engineering in the future.

Acknowledgements Part of this work was conducted at the Advanced Characterization Nanotechnology Platform of the University of Tokyo, supported by the "Nanotechnology Platform" of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. We are grateful to Hajime Hoshi (The University of Tokyo) for help with the initial negative stain experiment. We also thank Naoko Akimoto (WDB Co., Ltd.) for her technical assistance. In addition, we thank Moemi Tatsumi (Ajinomoto Co., Inc.) for her valuable advice. Finally, we are grateful to Kazuya Onomichi and Hiroshi Miyano (Ajinomoto Co., Inc.) for supporting this research. This work was also supported by Grants-in-Aid for Scientific Research (S), CREST from JST, the Japan New Energy and Industrial Technology Development Organization (NEDO), and the Japan Agency for Medical Research and Development (AMED) (to Y.F.); and Challenging Research (Exploratory), and the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP18am0101118j0002 (to K.T.). We thank James Allen, DPhil, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

Abrahamson MJ, Vazquez-Figueroa E, Woodall NB, Moore JC, Bommarius AS (2012) Development of an amine dehydrogenase for synthesis of chiral amines. Angew Chem Int Ed Engl 51(16):3969–3972. https://doi.org/10.1002/anie.201107813

Bai XC, Rajendra E, Yang G, Shi Y, Scheres SH (2015) Sampling the conformational space of the catalytic subunit of human gamma-secretase. Elife 4. https://doi.org/10.7554/eLife.11182

Baker PJ, Turnbull AP, Sedelnikova SE, Stillman TJ, Rice DW (1995) A role for quaternary structure in the substrate specificity of leucine dehydrogenase. Structure 3(7):693–705

Beckett PR, Hardin DS, Davis TA, Nguyen HV, Wray-Cahen D, Copeland KC (1996) Spectrophometric assay for measuring branched-chain amino acid concentrations: application for measuring

the sensitivity of protein metabolism to insulin. Anal Biochem 240(1):48–53. https://doi.org/10.1006/abio.1996.0329

Campollo O, Sprengers D, McIntyre N (1992) The BCAA/AAA ratio of plasma amino acids in three different groups of cirrhotics. Rev Invest Clin 44(4):513–518

Cao LC, Chen R, Xie W, Liu YH (2015) Enhancing the thermostability of feruloyl esterase EstF27 by directed evolution and the underlying structural basis. J Agric Food Chem 63(37):8225–8233. https://doi.org/10.1021/acs.jafc.5b03424

Fujiyoshi Y (1998) The structural study of membrane proteins by electron crystallography. Adv Biophys 35:25–80

Fujiyoshi Y, Mizusaki T, Morikawa K, Yamagishi H, Aoki Y, Kihara H, Harada Y (1991) Development of a superfluid helium stage for high-resolution electron microscopy. Ultramicroscopy 38(3):241–251. https://doi.org/10.1016/0304-3991(91)90159-4

Gibson T (1999) Biosensors: the stabilité problem. 27. https://doi.org/10.1051/analusis:1999270630

Gu KF, Chang TM (1990a) Conversion of ammonia or urea into essential amino acids, L-leucine, L-valine, and L-isoleucine, using artificial cells containing an immobilized multienzyme system and dextran-NAD+. 2. Yeast alcohol dehydrogenase for coenzyme recycling. Biotechnol Appl Biochem 12(3):227–236

Gu KF, Chang TM (1990b) Production of essential L-branched-chain amino acids in bioreactors containing artificial cells immobilized multienzyme systems and dextran-NAD+. Biotechnol Bioeng 36(3):263–269. https://doi.org/10.1002/bit.260360308

Haney PJ, Stees M, Konisky J (1999) Analysis of thermal stabilizing interactions in mesophilic and thermophilic adenylate kinases from the genus Methanococcus. J Biol Chem 274(40):28453–28458

Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J Physiol (Lond) 40:4–7

Huang T, Warsinke A, Kuwana T, Scheller FW (1998) Determination of L-phenylalanine based on an NADH-detecting biosensor. Anal Chem 70(5):991–997

Kataoka K, Tanizawa K (2003) Alteration of substrate specificity of leucine dehydrogenase by site-directed mutagenesis. J Mol Catal B Enzym 23(2):299–309. https://doi.org/10.1016/S1381-1177(03)00093-6

Kirino H, Aoki M, Miho A, Yumiko H, Masayuki O, Akihiko Y, Takayoshi W, Tairo O (1994) Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile. Thermus Thermophilus. Eur J Biochem 220(1):275–281. https://doi.org/10.1111/j.1432-1033.1994.tb18623.x

MacMullen C, Fang J, Hsu BY, Kelly A, de Lonlay-Debeney P, Saudubray JM, Ganguly A, Smith TJ, Stanley CA (2001) Hyperinsulinism/hyperammonemia syndrome in children with regulatory mutations in the inhibitory guanosine triphosphate-binding domain of glutamate dehydrogenase. J Clin Endocrinol Metab 86(4):1782–1787. https://doi.org/10.1210/jcem.86.4.7414

Matsuyama T, Soda K, Fukui T, Tanizawa K (1992) Leucine dehydrogenase from Bacillus stearothermophilus: identification of active-site lysine by modification with pyridoxal phosphate. J Biochem 112(2):258–265

Nagata S, Tanizawa K, Esaki N, Sakamoto Y, Ohshima T, Tanaka H, Soda K (1988) Gene cloning and sequence determination of leucine dehydrogenase from Bacillus stearothermophilus and structural comparison with other NAD(P)+-dependent dehydrogenases. Biochemistry (Mosc) 27(25):9056–9062

Nakamura K, Fujii T, Kato Y, Asano Y, Cooper AJ (1996) Quantitation of L-amino acids by substrate recycling between an aminotransferase and a dehydrogenase: application to the determination of L-phenylalanine in human blood. Anal Biochem 234(1):19–22. https://doi.org/10.1006/abio.1996.0043

Naruse H, Ohashi YY, Tsuji A, Maeda M, Nakamura K, Fujii T, Yamaguchi A, Matsumoto M, Shibata M (1992) A method of PKU screening using phenylalanine dehydrogenase and microplate system. Screening 1(1):63–66. https://doi.org/10.1016/0925-6164(92)90030-9

Ohshima T, Misono H, Soda K (1978) Properties of crystalline leucine dehydrogenase from Bacillus sphaericus. J Biol Chem 253(16):5719–5725

Ohshima T, Nagata S, Soda K (1985) Purification and characterization of thermostable leucine dehydrogenase from Bacillus stearothermophilus. Arch Microbiol 141(4):407–411. https://doi.org/10.1007/bf00428857

Rivero A, Allué Juan A, Grijalba A, Palacios M, Merlo Sergio G (2000) Comparison of two different methods for measurement of phenylalanine in dried blood spots. Clin Chem Lab Med 38. https://doi.org/10.1515/cclm.2000.110

Roh SH, Hryc CF, Jeong HH, Fei X, Jakana J, Lorimer GH, Chiu W (2017) Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM. Proc Natl Acad Sci U S a 114(31):8259–8264. https://doi.org/10.1073/pnas.1704725114

Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol 333(4):721–745

Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180(3):519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Sekimoto T, Fukui T, Tanizawa K (1994) Role of the conserved glycyl residues located at the active site of leucine dehydrogenase from Bacillus stearothermophilus. J Biochem 116(1):176–182

Stanley CA, Lieu YK, Hsu BY, Burlina AB, Greenberg CR, Hopwood NJ, Perlman K, Rich BH, Zammarchi E, Poncz M (1998) Hyperinsulinism and hyperammonemia in infants with regulatory mutations of the glutamate dehydrogenase gene. N Engl J Med 338(19):1352–1357. https://doi.org/10.1056/nejm199805073381904

Stanley CA, Fang J, Kutyna K, Hsu BY, Ming JE, Glaser B, Poncz M (2000) Molecular basis and characterization of the hyperinsulinism/hyperammonemia syndrome: predominance of mutations in exons 11 and 12 of the glutamate dehydrogenase gene. HI/HA Contributing Investigators. Diab 49(4):667–673

Suzuki K, Koizumi K, Ichimura H, Oka S, Takada H, Kuwayama H (2008) Measurement of serum branched-chain amino acids to tyrosine ratio level is useful in a prediction of a change of serum albumin level in chronic liver disease. Hepatol Res 38(3):267–272. https://doi.org/10.1111/j.1872-034X.2007.00268.x

Tachibana S, Suzuki M, Asano Y (2006) Application of an enzyme chip to the microquantification of l-phenylalanine. Anal Biochem 359(1):72–78. https://doi.org/10.1016/j.ab.2006.09.006

Tajiri K, Shimizu Y (2013) Branched-chain amino acids in liver diseases. World J Gastroenterol 19(43):7620–7629

Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65(1):1–43. https://doi.org/10.1128/mmbr.65.1.1-43.2001

Wagner A, Fell DA (2001) The small world inside large metabolic networks. Proc Biol Sci 268(1478):1803–1810. https://doi.org/10.1098/rspb.2001.1711

Yamaguchi H, Kamegawa A, Nakata K, Kashiwagi T, Mizukoshi T, Fujiyoshi Y, Tani K (2019) Structural insights into thermostabilization of leucine dehydrogenase from its atomic structure by cryo-electron microscopy. J Struct Biol 205(1):11–21. https://doi.org/10.1016/j.jsb.2018.12.001

Yokozeki K (1988) Kouso-hou niyoru kougaku kassei aminosan no seisan [Production of optical activity by enzymatic method]. J of Brew Soc Japan 83(4):230–237. https://doi.org/10.6013/jbrewsocjapan1988.83.230

Zhao Y, Wakamatsu T, Doi K, Sakuraba H, Ohshima T (2012) A psychrophilic leucine dehydrogenase from Sporosarcina psychrophila: Purification, characterization, gene sequencing and crystal structure analysis. J Mol Catal B: Enzym 83:65–72. https://doi.org/10.1016/j.molcatb.2012.06.018

Chapter 11
Structure, Function and Physiology of 5-Hydroxytryptamine Receptors Subtype 3

Eric Gibbs and Sudha Chakrapani

Abstract 5-hydroxytryptamine receptor subtype 3 (5-HT_3R) is a pentameric ligand-gated ion channel (pLGIC) involved in neuronal signaling. It is best known for its prominent role in gut-CNS signaling though there is growing interest in its other functions, particularly in modulating non-serotonergic synaptic activity. Recent advances in structural biology have provided mechanistic understanding of 5-HT_3R function and present new opportunities for the field. This chapter gives a broad overview of 5-HT_3R from a physiological and structural perspective and then discusses the specific details of ion permeation, ligand binding and allosteric coupling between these two events. Biochemical evidence is summarized and placed within a physiological context. This perspective underscores the progress that has been made as well as outstanding challenges and opportunities for future 5-HT_3R research.

Keywords 5-HT · Serotonin · 5-HT_3R · Pentameric ligand-gated ion channel · Neurotransmission · Membrane biophysics

Introduction

5-hydroxytryptamine (5-HT) or serotonin is a versatile neurotransmitter whose functions range from visceral reactions to higher-order neuronal processes. 5-HT affects cellular processes by binding to a large family of membrane-bound 5-HT receptors (5-HTR). 5-HTRs are mostly G-protein coupled receptors (GPCR), and act through

E. Gibbs (✉) · S. Chakrapani (✉)
Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, OH 44106-4970, USA
e-mail: edg42@case.edu

S. Chakrapani
e-mail: sxc584@case.edu

S. Chakrapani
Department of Neuroscience, School of Medicine, Case Western Reserve University, Cleveland, OH 44106-4970, USA

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_11

relatively slow metabotropic processes. The sole exception is 5-HT_3R, which is a pLGIC that participates in fast neurotransmission (Lummis 2012).

5-HT_3Rs are expressed throughout the body and are especially known for their specialized roles in brain-gut circuitry. In this context, the receptors regulate gut motility, secretion, emetic reflux and visceral perception and are important therapeutic targets for chemotherapy-induced vomiting and irritable bowel syndrome (IBS). Elsewhere in the brain, 5-HT_3Rs play a mostly modulatory role in synaptic signaling and are potential targets of several psychiatric disorders (Rojas and Slusher 2012). 5-HT_3Rs share a quaternary structure with three other vertebrate pLGICs, nicotinic acetylcholine receptors (nAChR), γ-amino butyric acid A receptors ($GABA_AR$) and glycine receptors (GlyR). Together, these ion channels are integral components of neurotransmission with significant interplay. Therefore, the underlying mechanisms of 5-HT_3R function and modulation are not only relevant to 5-HT signaling but general neuronal communication.

5-HT_3Rs were first recognized as ion channels in 1989 (Derkach et al. 1989). In 1991, the first of five 5-HT_3R genes, 5-HT_{3A}, was discovered and the amino acid sequence revealed that 5-HT_3Rs are members of the pLGIC family (Maricq et al. 1991). pLGICs assemble as a complex of five homologous subunits that are often heterogeneous. To date, there are five 5-HT_3R genes termed A-E (Niesler et al. 2003; Davies et al. 1999). Only 5-HT_{3A} can assemble as a functional homopentamer, 5-HT_{3A}R. (5-HT_{3A} refers to a single subunit and 5-HT_{3A}R refers to an assembled pentamer). 5-HT_{3A} is an obligate participant in all other 5-HT_3R complexes and the most well-studied complexes are 5-HT_{3A}R and heteromeric 5-$HT_{3A/B}$R.

The first section of this chapter is a general overview of the physiological role of 5-HT_3R and its general structure. The physiological role of 5-HT_3R requires three fundamental properties. (1) 5-HT_3R selectively conducts and gates ionic flux across the membrane, (2) 5-HT_3R strongly and specifically binds 5-HT and (3) 5-HT_3R allosterically connects channel activity to ligand binding. Each of these properties is presented in its own section that discusses its physiological implications, underlying mechanisms and regulation by endogenous and exogenous ligands. The chapter concludes with a perspective on outstanding questions and future directions of 5-HT_3R research.

General Overview

Expression and Physiological Role

5-HT3Rs in the PNS

5-HT_3Rs participate most prominently in CNS-GI signaling, consistent with the fact that more than 90% of 5-HT is found in the gut. Specifically, 5-HT_3Rs are found on nodose ganglion afferent neurons that innervate the GI tract (Leal-Cardoso et al.

1993). The receptors are found at multiple locations on nodal afferents including, dendrites in the gut, cell bodies at the ganglion (located at the base of the skull) and presynaptic terminals in the brain stem (Powley et al. 2011). Dendritic 5-HT_3Rs directly respond to 5-HT released by enterochromaffin cells during digestion. (Hillsley et al. 1998; Hillsley and Grundy 1998). Receptors at the ganglion and brainstem can be activated by 5-HT circulating in the bloodstream because the barrier between them and the bloodstream is leaky (Lacolley et al. 2006a,2006b; Baptista et al. 2007). Once activated, 5-HT signals are transmitted to the brainstem via glutamatergic synapses in the nucleus tractus solitarius (NTS). There, 5-HT_3Rs are found presynaptically where they modulate glutamate release, and there is at least one report of postsynapatic 5-HT_3Rs that shape the NTS response (Glaum et al. 1992; Jeggo et al. 2005). Nodal afferent stimulation is summarized in Fig. 11.1.

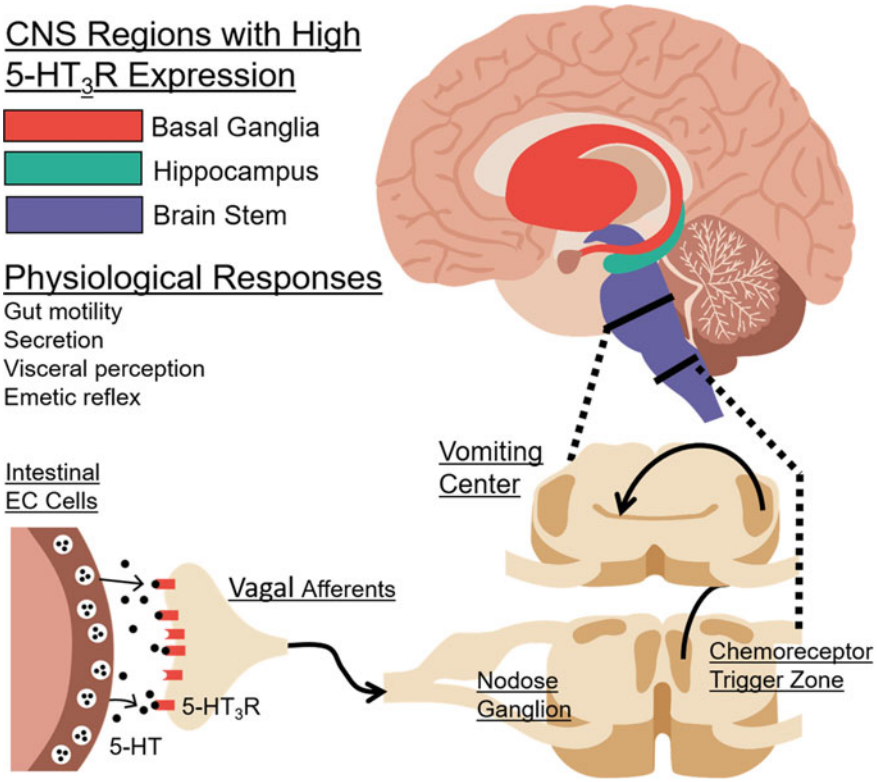


Fig. 11.1 Localization and physiological signaling of 5-HT_3R. 5-HT released from enterochromaffin cells in the intestine stimulates 5-HT_3Rs on the nerve terminals of vagal afferents. In addition, circulating 5-HT can stimulate 5-HT_3Rs in the nodose ganglion or within the brain stem. These initiate signals within the brain that contribute to the listed physiological responses. 5-HT_3Rs are also present in other CNS regions where they play a mostly modulatory role at glutamatergic, GABAergic and dopaminergic synapses

5-HT_3Rs on vagal nerves are not required for enteric motility but are important in modulating motility according to environmental factors (Keating and Spencer 2010; Blackshaw and Grundy 1993). For example, in the gut, glucose stimulates enterochromaffin cells which in turn release 5-HT. 5-HT then activates 5-HT_3Rs that inhibit gastric motility, allowing time for digestion (Raybould et al. 2003). Glucose in the bloodstream can further impact nodal afferents by directly increasing excitability and by increasing surface expression of 5-HT_3Rs throughout the cell membrane and presynaptic terminals (Grabauskas et al. 2010; Babic et al. 2012; Wan and Browning 2008). The influence of glucose on 5-HT_3R signaling under both normal and pathophysiological conditions requires further investigation. In addition to nutrient modulation, antigens also indirectly activate 5-HT_3Rs via enterochromaffin cells. Antigen-induced 5-HT_3R stimulation leads to surface expression of immunogenic NK1 receptors on afferent vagal neurons, implicating 5-HT_3Rs in long-term nausea, antigen-induced pain and inflammation (Moore et al. 2002). 5-HT_3R antagonists have been used effectively as anti-emetics during cancer treatment and more recently for treatment of IBS (Zheng et al. 2017; Rojas and Slusher 2012). This is discussed further in the section on ligand binding.

Though peripheral 5-HT_3R input is generally associated with vagal afferents, 5-HT_3Rs are also found in other neurons though there are significant variations between species. In some species, 5-HT_3Rs have been found in the enteric nervous system, spinal nociceptive neurons and in the trigeminal nucleus, which innervates the jaw and face (Parker et al. 1996; Bertrand et al. 2000; Glaum et al. 1990). 5-HT_3Rs are also found in non-excitable cells within the GI tract including interstitial cells of Cajal, which act as pacemaker cells for smooth muscle, and enteroendocrine cells, though not enterochromaffin cells (Glatzle et al. 2002).

5-HT_3Rs in the CNS

Radioligand binding and mRNA expression show that 5-HT_3Rs are found in several brain structures. In the brainstem, 5-HT_3Rs are found on structures that integrate inputs from other sensory neuronal circuits (Parker et al. 1996; Miyake et al. 1995). 5-HT_3R expression is particularly high in the NTS and area postrema. The receptor is also present in the adjacent dorsal motor nucleus of the vagus (DMV), implicating 5-HT_3R activity in physiological responses such as vomiting and blood pressure regulation (Babic and Browning 2014; Sevoz et al. 1997; Mussa et al. 2008).

In humans, 5-HT_3Rs are also detected at relatively high levels in the hippocampus and basal ganglia and at lesser levels in the cortex (Bufton et al. 1993; Parker et al. 1996). In these regions, there are some serotonergic synapses with postsynaptic 5-HT_3Rs, but 5-HT_3Rs are especially found presynaptically where they modulate GABAergic and dopaminergic synapses (Sugita et al. 1992; Mylecharane 1995; Puig et al. 2004). In particular, 5-HT_{3A}R homomers have conductance, ligand binding and kinetic properties tailored to this role, as will be discussed throughout the chapter. Understanding the effects of this modulation is difficult as there are significant variations in 5-HT_3R expression and function between species (Miyake et al. 1995).

5-HT_3R modulation is implicated in complicated neurobehaviors such as depression, anxiety, bipolar disorder, pain perception and appetite regulation (Fakhfouri et al. 2019; Rajkumar and Mahesh 2010; Wu et al. 2012; Jones et al. 1988; Liang et al. 2011; Thompson et al. 2006b). 5-HT_3Rs are attractive drug targets as they exhibit less adverse side effects and use dependence than traditional drugs that directly target synaptic activity. 5-HT_3R antagonists have been used successfully to treat schizophrenia and alcohol abuse and several studies suggest potential in other psychiatric disorders (Dorostkar and Boehm 2007).

Differential Expression of 5-HT_3 Subunits and Isoforms

Expression patterns of 5-HT_3R are further complicated when considering different isoforms and subunit types. As will be discussed through the chapter, the properties of 5-HT_{3A}Rs are different to 5-$HT_{3A/B}$Rs and expression of one over the other has physiological implications. In humans, 5-HT_{3A} and 5-HT_{3B} are expressed in near equal ratios in the small intestine but 5-HT_{3B} is expressed at much higher levels in the brain (Tzvetkov et al. 2007). Relative expression also varies between different brain structures. For example, 5-HT_{3B} transcription is six times higher than 5-HT_{3A} in the amygdala while the ratio is roughly equal in the hippocampus (Brady et al. 2007; Tzvetkov et al. 2007). Surface expression of 5-HT_{3A}R homomers and 5-$HT_{3A/B}$R heteromers is also dependent on chaperone proteins and glycosylation (Boyd et al. 2002). In particular, RIC-3 specifically promotes 5-HT_{3A}R surface expression over 5-$HT_{3A/B}$R (Cheng et al. 2007).

Similarly, the 5-HT_{3C}, 5-HT_{3D} and 5-HT_{3E} subunits all form functional channels with the 5-HT_{3A} subunit (Niesler et al. 2003). These subunits are expressed in peripheral tissues, particularly in the colon, and may be expressed in the brain as well (Niesler et al. 2007; Kapeller et al. 2011). The study of these subunits is limited by the lack of subunit-specific antibodies and their physiological function is largely unknown. For simplicity subunits C-E will not be discussed in further detail.

5-HT_3 subunit isoforms also display tissue-specific expression. In humans, 5-HT_{3A} isoforms have different localization patterns and functional properties (Brüss et al. 2000). 5-HT_{3A} isoform expression also shifts during development in mouse cell lines (Emerit et al. 1995). 5-HT_{3B} isoforms transcribed in the brain are not found in peripheral tissue, and *visa-versa* (Tzvetkov et al. 2007; Jensen et al. 2008). The major 5-HT_{3B} isoform in either tissue encodes the same protein but the isoforms differ in peptide signaling sequences. A minor yet significant portion of brain isoforms are truncated up to and including the β1-β2 loop. Overall, isoform expression appears tissue and development specific and thus physiologically important.

Basic Structural Topology

Early pLGIC studies identified amino acids important for specific functions and clues about the underlying structure. However, until recently, direct structural information had been elusive as eukaryotic pLGICs are not easily crystalized. Prior to the last five years, pLGIC structural studies were mostly limited to bacterial homologues, the naturally enriched torpedo nAChR or acetylcholine binding protein (AChBP) (Unwin 2005; Brejc et al. 2002; Kesters et al. 2013; Hilf and Dutzler 2008). In 2014, mouse 5-HT_{3A}R was finally able to be crystallized in the presence of stabilizing nanobodies (Hassaine et al. 2014). There has been a recent explosion of pLGIC structural data with major technological advancements in cryo-electron microscopy (cryo-EM). These include several 5-HT_{3A}R structures bound to different ligands and in multiple conformational states (Basak et al. 2018a, b; 2019; Polovinkin et al. 2018; Basak et al. 2020; Zarkadas et al. 2020). Figure 11.2 shows the fully-assembled structure of 5-HT_{3A}R and Fig. 11.3 highlights the historical nomenclature of different regions. Throughout the chapter, residue numbering is consistent with that of the apo mouse 5-HT_{3A}R structure published by Basak and colleagues (PDB 6BE1).

pLGICs assemble as five homologous subunits about a pseudo-symmetric five-fold axis as shown in Fig. 11.2. The subunit matching is not fully understood, but important factors include complementary shape and charge at subunit interfaces, subunit-specific glycosylation patterns, specific chaperone proteins and spatial constraints in closing the five-fold ring structure (Walsh et al. 2018; Boyd et al. 2002; Phulera et al. 2018). It is generally believed that unmatched subunits are retained in the ER and eventually degraded (Davies et al. 1999). Specific to 5-HT_3R, functional expression requires at least one 5-HT_{3A} subunit, but the stoichiometry of heteromeric receptors is unknown. The stoichiometry has implications for ligand binding and is discussed further in that section.

Individual subunits of pLGICs can be divided into three domains, the extracellular domain (ECD), transmembrane domain (TMD) and intracellular domain (ICD). The structure of the ECD is remarkably well-conserved across pLGICs (Brejc et al. 2002). It consists of an N-terminal alpha helix followed by ten beta strands forming a beta-barrel structure. Clusters of residues important for ligand binding are named loops 'A-F' though it was later determined that not all of these clusters are technically loops (Brejc et al. 2002). Another set of loops at the ECD-TMD interface are important for coupling ligand binding to channel gating. These include the β1-β2 loop, the β6-β7 loop (also known as the cys-loop), the β8-β9 loop and the covalent link between β10 and the TMD.

The TMD has consists of 4 bundled membrane-spanning alpha helices named M1-M4. The M2 helix forms the channel pore and is flanked by M3 counterclockwise and M1 clockwise when viewed from the ECD. M4 is most exposed to the membrane environment and acts as a lipid sensor (Hénault et al. 2015). The loops between the TM helices are also important for channel function. The extracellular M2-M3 linker transduces movements at the ECD-TMD interface to the channel pore. The ICD is comprised of the intracellular loops between the M1 and M2 helices and between

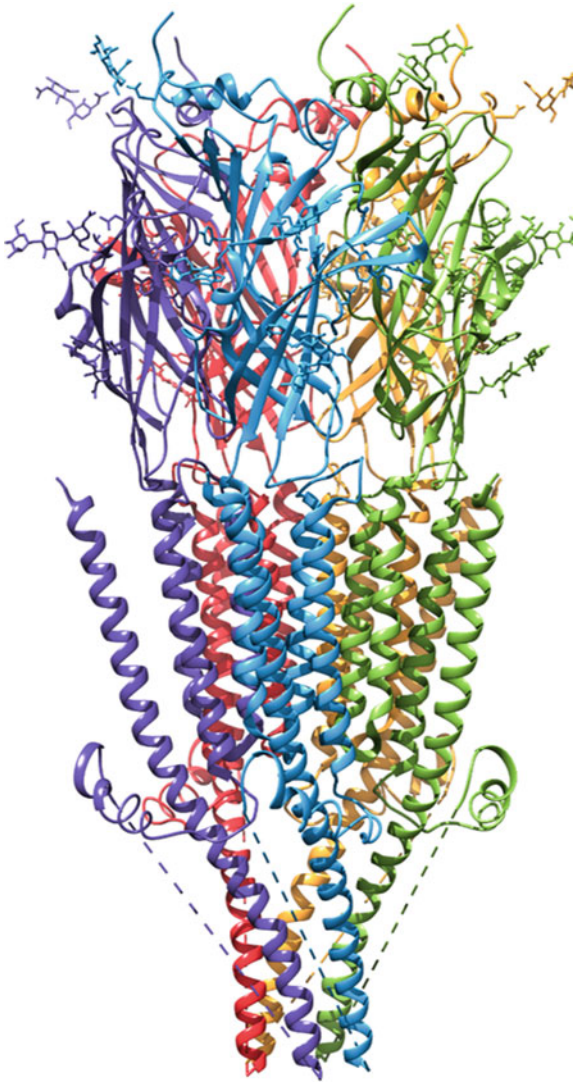


Fig. 11.2 Full pentameric structure of mouse 5-HT_{3A}R (PDB 6NP0) (Basak et al. 2019)

the M3 and M4 helices. The M1-M2 linker is relatively short but plays a role in ionic charge selectivity and desensitization. The M3-M4 linker is 70–120 amino acids long, has the most heterogeneity among pLGICs and includes a large stretch of structurally disordered residues (Basak et al. 2018b). It has an important role in membrane trafficking, regulating channel conductance and interactions with other signaling proteins. The two structured elements of the ICD are the MX helix, which lies parallel to the membrane shortly following the M3 helix, and the MA helix, which directly precedes the M4 helix.

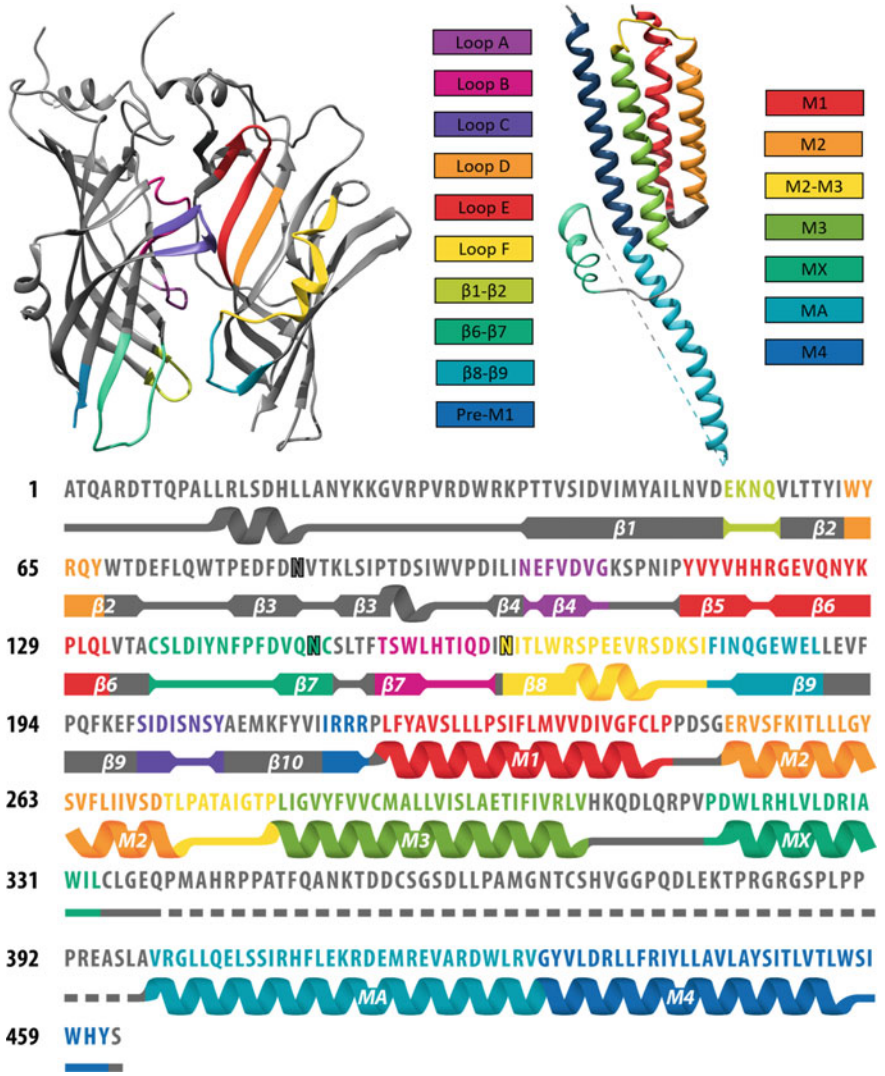


Fig. 11.3 Labeled sequence and 3D-structure of 5-HT_{3A} subunit components (PDB 6NP0) (Basak et al. 2019). Top left, two neighboring subunit ECDs are shown. Binding site Loops A-F and ECD-TMD interface regions are color-coded by historical nomenclature. Top right, the TMD and ICD of a single subunit are similarly colored. Note that the colors are repeated from the ECD image. Below, the amino-acid sequence of 5-HT_{3A} color coded to match the above images. The break between the ECD and TMD-ICD image is between β10 and M1. Glycosylated asparagines are outlined

Ion Permeation and Channel Gating

5-HT_3R was first recognized as an ion channel from electrophysiology studies and characterized as a pLGIC from its amino acid sequence (Maricq et al. 1991; Derkach et al. 1989). As with all pLGICs, the M2 helix forms the channel pore, but elements of the ECD and ICD also affect ion permeation (Schofield et al. 1987). This section describes observations of 5-HT_3R ion permeation and their physiological implications. It then addresses the underlying physical mechanisms of these observations. The section closes with a brief summary of antagonists that act at the channel pore.

Observed Properties of 5-HT_3R Ion Conductance

Early electrophysiology studies agreed that 5-HT_3Rs are cationic with a similar relative conductance for Na^+/K^+ with a reversal potential near 0 mV, but other properties diverged significantly (Gill et al. 1995; Derkach et al. 1989; Yang 1990). Some studies reported 5-HT_3Rs had sub-picosiemen single channel conductance, were calcium permeable and displayed strong voltage rectification (Gill et al. 1995; Yang et al. 1992; Brown et al. 1998). Other studies reported a channel with much a higher single channel conductance (9–15 pS), reduced divalent cation permeability and no intrinsic rectification (Hussy et al. 1994; Derkach et al. 1989; Jones and Surprenant 1994). Discovery of the 5-HT_{3B} subunit confirmed that most of these differences could be attributed to channel subunit composition (Davies et al. 1999; Dubin et al. 1999; Hapfelmeier et al. 2003). Specifically, 5-HT_{3A}R homomers fit the first description and 5-$HT_{3A/B}$R heteromers fit the latter. Interestingly, the M2 portions of these channels are quite similar suggesting that elements outside the pore helix regulate conductance properties.

Differences in channel conductance and permeability have important implications in neuronal communication. For example, presynaptic 5-HT_3Rs modulate inhibitory GABA release in a calcium-dependent manner (Koyama et al. 2000; Rondé and Nichols 1998; Choi et al. 2007). This suggests that these presynaptic receptors are 5-HT_{3A}R homomers as only 5-HT_{3A}Rs are calcium permeable. Another example, the high conductance of 5-$HT_{3A/B}$R will lead to larger changes in membrane potential upon activation. A common theme of 5-HT_3R is that 5-HT_{3A}Rs seem better suited for presynaptic modulation and 5-$HT_{3A/B}$Rs seem better for postsynaptic transmission. However, mRNA expression and western analysis suggest that the roles of 5-HT_{3A}R and 5-$HT_{3A/B}$R in vivo are more subtle and vary across neuronal populations (Férézou et al. 2002; Brady et al. 2007).

In addition to fast synaptic signaling, 5-HT_{3A}R calcium activity has long-term effects. 5-HT_{3A}R calcium flux is prolonged by 5-HT open-pore block at higher 5-HT concentrations, which does not occur in 5-$HT_{3A/B}$Rs (Hapfelmeier et al. 2003; Corradi et al. 2009). Prolonged calcium signaling is associated with synaptic plasticity, suggesting 5-HT_{3A}R may be important in memory and development. Indeed,

5-HT_{3A} expression is relatively high in the hippocampus compared to other regions of the brain and 5-HT_3R expression is important in the development and pathology of rat pyramidal and Purkinje cells (Brady et al. 2007; Oostland et al. 2013; Rosenberg et al. 1997; Roerig et al. 1997; Zhang et al. 2011; Choi et al. 2007). Both 5-HT_{3A}R and 5-$HT_{3A/B}$R are blocked by elevated extracellular calcium. Elevated extracellular calcium occurs near dying cells and calcium block can protect further damage by overstimulation following injury.

The distinct properties of 5-HT_{3A}R and 5-$HT_{3A/B}$R are relatively well known, but their physiological roles are not yet understood. This is partially due to the complex interplay between channel subtypes and other neurotransmitter signaling, but also due to the lack of quality antibodies to distinguish between the two populations. This technical challenge is further exasperated in the C, D and E 5-HT_3 subunits. Future research will hopefully reveal how subunit-specific channel properties influence the diverse physiological roles of 5-HT_3R.

Mechanisms Underlying Ion Permeation

Physical Description of the Permeation Pathway

Unlike many selective ion channels, pLGICs do not have a unique well-defined ion selectivity filter. Rather, channel properties are influenced by spatially dispersed elements in the ECD, TMD and ICD as shown in Fig. 11.4. The extracellular conductance pathway consists of a bowl-shaped vestibule with an approximate average inner-diameter of 20 Ångstroms. It has been proposed that ions can enter this vestibule either through the top and/or from gaps between the subunits and that glycosylation sites help define the conduction pathway for some pLGICs (Miller and Aricescu 2014; Phulera et al. 2018; Zhu et al. 2018). 5-HT_{3A}R structures show several charges along the inner vestibule wall, including D91 and D105 (Basak et al. 2018b; Hassaine et al. 2014). On the background of a high conductance 5-HT_{3A}R mutant (with mutated intracellular arginines as discussed below), further mutation of D91 or especially D105 reduced channel conductance, calcium permeability and inhibition by elevated extracellular calcium (Livesey et al. 2011). Molecular dynamic simulations of nAChR show that these residues coordinate positively charged cations and there is an elevated cation concentration throughout the extracellular vestibule (Wang et al. 2008). Residues corresponding to D105 are conserved in other cationic pLGICs while anionic pLGICs have positively charged residues at this position, suggesting extracellular charge selection is a conserved mechanism across the family (Hansen et al. 2008).

D271 or D20′ (The prime notation is the position relative to the cytoplasmic end of M2), is a pore facing negatively charged residue at the ECD-TMD interface. This is the most extracellular of three ‘rings of charge’ that were identified as early determinants of cation conductance and selectivity in nAChRs (Miyake et al. 1995; Imoto et al. 1988). D20′ in 5-HT_{3A} is conserved across species and in some 5-HT_{3B}

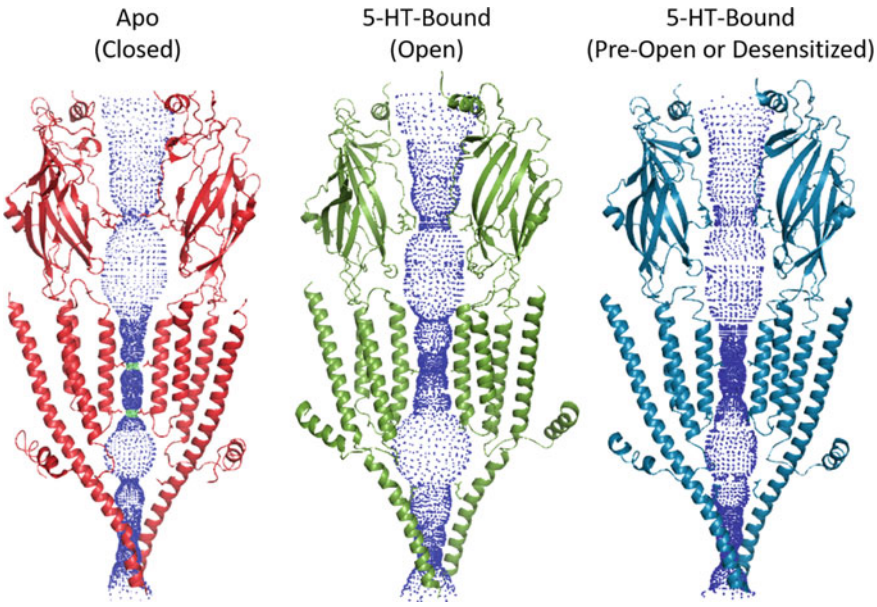


Fig. 11.4 Channel permeation pathway in different conformations (PDBs 6BE1, 6DG7 and 6DG8) (Basak et al. 2018a; Basak et al. 2018b). The L9′ and E-1′ gates are shown most clearly in the apo structure near the constriction points shown in green. The residues above L9′ form a hydrophobic barrier that requires ion solvation for passage. The M2 helices splay outwards and rotate in the 5-HT-bound (open) state compared to the apo (closed) state, widening the channel pore. The 5-HT-bound (pre-open or desensitized) state show an intermediate M2 helix position where the E-1′ gate is the most constricted point. MD simulations suggest that this state is mostly impermeable

subunits as well, though in human 5-HT_{3B} it is an asparagine. Homomeric receptors with a 5-HT_{3A} mutant D20′A do not differ in monovalent cation selectivity but have a lower channel conductance and altered kinetics (Livesey et al. 2008; Hu and Lovinger 2005). This could occur because anions are already disfavored within the extracellular vestibule but mutations to D20′ disrupt cation flow. The D20′A mutant also shows reduced calcium permeability, possibly due to weakened electrostatic interactions at the pore entry. N20′ in the 5-HT_{3B} subunit may contribute to the calcium impermeability of 5-$HT_{3A/B}$Rs.

E250 or E-1′ forms the second ring of negative charge in 5-HT_{3A}R at the cytosolic end of the M2 helix. This glutamate is conserved across several nAChR subunits. A non-polar mutation of E-1′A results in a channel that shows no monovalent charge selectivity and charge selection can be reversed when combined with other mutations (Thompson and Lummis 2003; Gunthorpe and Lummis 2001). Recent structures depicting 5-HT_{3A}R in an open conformation show that the channel pore is wider at the extracellular end than at the cytoplasmic end as seen in Fig. 11.4. This supports evidence that E-1′ is a stronger determinant of charge and size selectivity than D20′ (Basak et al. 2018a; McKinnon et al. 2011; Polovinkin et al. 2018). However, 5-HT_{3B} subunits and some nAChR subunits do not contain a negatively charged residue in

the E-1′ position and monovalent cation selectivity is not dramatically altered (Peters et al. 2005). Thus negative charge is not required in every subunit within the pentamer for cation selection. The overall charge reduction may however contribute to calcium impermeability or lack of voltage rectification in 5-$HT_{3A/B}$R (Gunthorpe and Lummis 2001).

The pore lining residues of M2 lie between E-1′ and D20′. L260 or L9′ is at the center of the pore and the narrowest part of the pore in 5-HT_{3A}R apo and antagonist-bound structures. The position of L9′ can be seen in the Apo (Closed) pore profile in Fig. 11.4. L9′, V13′ and I17′ form a hydrophobic barrier within the channel pore that is generally conserved across pLGICs (Corry 2006; Revah et al. 1991; Filatov and White 1995). Mutations to L9′ and V13′ affect open channel stability (Dang et al. 2000). Cysteine labelling studies suggest that these residues form a vestibule large enough to accommodate solvated ions and MD simulations show ions must be solvated to cross this barrier (Reeves et al. 2001). In 5-HT_{3A}Rs, a pair of polar residues, S2′ and T6′ are intracellular to the hydrophobic barrier. However, polar residues in these positions are not conserved across pLGICs, including the 5-HT_{3B} subunit, which has an alanine at the 2′ position (Davies et al. 1999). The uncharged alanine may however, once again, contribute to specific properties of 5-$HT_{3A/B}$Rs (Das and Dillon 2003).

The 5-HT_3 structure shows that the 5 MA helices coil and form an inverted pyramid that occludes intracellular ion permeation along the pore axis but ions can pass through the pore through intracellular portals that line the MA helix (Kelley et al. 2003; Basak et al. 2018a; Deeb et al. 2007). Structural and mutational studies show this is a common feature of pLGICs, though the MA helix is unstructured in anionic pLGIC structures (Carland et al. 2009; Zhu et al. 2018; Phulera et al. 2018; Du et al. 2015; Everitt et al. 2009). One key piece of evidence for ion permeation via intracellular portals is the remarkably low single channel conductance of 5-HT_{3A}R relative to 5-$HT_{3A/B}$R and other pLGICs. This is particularly surprising as 5-$HT_{3A/B}$R contains hydrophobic residues within the channel pore. It was found that three arginines within the MA helix, R416, R420 and R424, were responsible for the low conductance of 5-HT_{3A}R. Mutation of these residues to their 5-HT_{3B} counterparts, R416Q, R420D and R424A gave 5-HT_{3A}R a single channel conductance comparable to 5-$HT_{3A/B}$R (Kelley et al. 2003; Hales et al. 2006). Cysteine mutant labelling, MD simulations and 5-HT_{3A}R structures show that these arginines line the intracellular portals (Basak et al. 2018a; Deeb et al. 2007). These structures also suggest that the MX helix may be positioned to act as another conformational gate in different functional states. Mutation of these arginines also eliminates phosphate-dependent inhibition at hyperpolarized membrane potentials and it is likely that negatively charged phosphates that block the channel are coordinated by arginine residues (Noam et al. 2008).

Conformational Changes to the Channel Pore with Different Functional States

In the extracellular permeation pathway, ligand binding leads to a counter-clockwise rotation of individual 5-HT_3 subunits and loop rearrangement (Basak et al. 2018a). This rotation exposes more negative charges to the vestibule particularly near the entrance to the TMD pore. There, the mobile β1-β2 loop, β6-β7 loop and M2-M3 linker form a strong electronegative pocket. The mobile loops are important for allosteric coupling, making it difficult to clearly define their effects on channel conductance, though the anionic glycine receptor has more neutrally charged residues at this interface (Du et al. 2015).

Within the membrane pore, the M2 helix undergoes significant changes between different functional states, which can be visualized in Fig. 11.4. Cysteine labelling and, more recently, ligand-bound structures of 5-HT_{3A}R have captured these different states (Reeves et al. 2001; Basak et al. 2018a; 2019; Panicker et al. 2002; Polovinkin et al. 2018). Assigning structures to specific functional states is challenging, but in general, the positioning of the M2 helix in apo and ligand-bound pLGIC structures is consistent with the expected functional state (Corrie and Baenziger 2013).

Apo 5-HT_{3A}R structures show the M2 helix mostly perpendicular to the membrane with two constriction points, the narrowest being L9′. MD simulations show that the L9′/V13′ hydrophobic barrier is impermeable to ions and water in this state. Apo 5-HT_{3A}R is also constricted at E-1′. Unlike the hydrophobic gate, this constriction remains closed in the desensitized state and E-1′ is referred to as the desensitization gate. Antagonist-bound structures show the M2 helices in a similar conformation to the apo state and it is reasonable to assign these structures to a closed channel state (Basak et al. 2019; Polovinkin et al. 2018).

Different M2 arrangements have been observed in serotonin-bound 5-HT_{3A}R structures (Basak et al. 2018a; Polovinkin et al. 2018). In one observed serotonin-bound state, the M2 helix is slightly kinked at the L9′ position and rotated clockwise relative to the apo state. Both extracellular and intracellular entrances to the TMD pore are splayed radially outward (Basak et al. 2018a). L9′ is also shifted away from the pore and slightly rotated. This tilt and rotation widen the pore, similar to a camera iris. Due to the kink, the desensitization gate is also slightly wider and the net effect is a pore wide enough to allow permeation of solvated ions. These structural observations are supported by open-state dependent labeling of mutated cysteines near the L9′ position and MD simulations. Thus one can assign these structures to open states (Panicker et al. 2002).

In another serotonin-bound state, the M2 helix is no longer kinked but is still tilted relative to the closed position (Basak et al. 2018a). The extracellular end of M2 is splayed outward, though not as much as the open state structure, and the intracellular end is now positioned radially inward. The L9′ barrier is still wide enough to allow ion passage like the open state, but the desensitization gate at E-1′ is even further constricted than the closed state. MD simulations show that permeation events are much rarer and the simulations do not include residues from the M3-M4 linker, which

may form an additional barrier to ion permeation in this state. It is unclear however whether this state is a pre-open or desensitized state (Corradi et al. 2009; Solt et al. 2007).

Pore Blocking Antagonists

Several 5-HT_3R antagonists bind within the pore region and block ion conduction including picrotoxin, bilobalide and ginkgolide B. In human 5-HT_3Rs, pore block by these compounds is eliminated by T6′S or S12′A mutations suggesting these residues may hydrogen bond with the antagonist (Thompson et al. 2011b). Interestingly, S2′A mutations enhance channel block and even cause ginkogolide B to become trapped in the pore upon channel closure, likely by hydrophobic contacts that further stabilize the blocker deeper in the pore. 5-$HT_{3A/B}$R and mouse 5-HT_{3A}R are 4–6 times less sensitive than human 5-HT_{3A}R to these antagonists, consistent with differences in pore lining residues (Das and Dillon 2005, 2003; Thompson et al. 2011b). In addition to natural compounds, tricyclic antidepressants such as fluoxetine and antipsychotic medications can act as pore blockers and this may contribute to their therapeutic effects (Gumilar and Bouzat 2008; Rammes et al. 2004).

Other compounds bind at both the orthosteric binding site as well as the channel pore, including the primary agonist 5-HT, the blood pressure medication diltiazem and morphine (Hapfelmeier et al. 2003; Thompson et al. 2011b; Baptista-Hon et al. 2012). The anti-malarial compound mefloquine also acts as a channel blocker while structurally similar anti-malarial compounds act at the orthosteric binding site (Thompson and Lummis 2008). The binding site for these compounds overlaps with that of picrotoxin but is slightly higher as diltiazem binding is affected by 7′ and 12′ but not 2′ or 6′ mutations. 5-HT_3R pore block may contribute to the side-effects of certain drugs and as previously mentioned prolongs 5-HT_{3A}R signaling. Improved understanding of pore block mechanisms could improve drugs by increasing the efficacy of anti-depressants or limiting off-target effects of other drugs.

Ligand Binding

Ligand binding is the initial signal that governs channel activity. This section first presents common 5-HT_3R agonists, antagonists and partial agonists that bind at the orthosteric binding site and their role in physiology or medicine. Then we look at specific interactions that mediate ligand binding and subsequent rearrangements within the binding pocket.

Description of the Ligand Binding Site and Orthosteric Ligands

The neurotransmitter binding site is located at subunit interfaces as shown in Fig. 11.5. 5-HT_{3A}R presents five identical binding sites, but not all of the subunit interfaces are equivalent in heteromeric channels. Considering only the 5-HT_{3A} and 5-HT_{3B} subunits, there are four possible interfaces, A(+)/A(−), A(+)/B(−), B(+)/A(−) and B(+)/B(−), where + and − refer to the principal and complementary subunits. There is strong pharmacological evidence that only the A(+)/A(−)

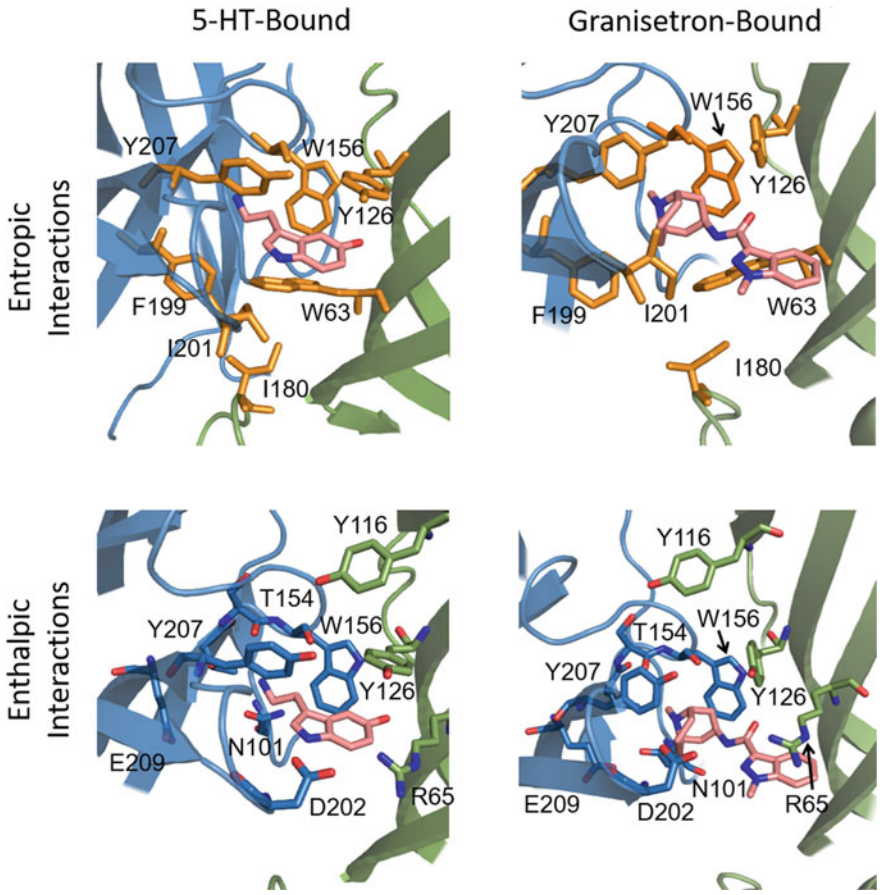


Fig. 11.5 Interactions within the binding pocket for both 5-HT and the competitive antagonist granisetron (PDBs 6DG7, 6NP0) (Basak et al. 2018a; Basak et al. 2019). Residues discussed within the text are shown and labelled. Top panels show a selection of residues involved in entropic interactions, shown in orange, and bottom panels show enthalpic interactions, colored by heteroatom. Note the tropane moiety of granisetron more completely fills the binding pocket, and how the elongated structure widens the gap between subunit interfaces

interface binds both agonists and competitive antagonists. Mutations at the binding interface that swap 5-HT_{3A} residues for those found in 5-HT_{3B} dramatically reduce the receptor's EC_{50} while 5-HT_{3B} to 5-HT_{3A} swaps have little effect (Lochner and Lummis 2010). Additionally, cysteine mutant labels and double cysteine mutations across the A(+)/A(−) interface ablate agonist activity but can be rescued with DTT, while this treatment at other possible binding interfaces has no effect (Thompson et al. 2011a). Similar results show granisetron, a competitive antagonist, only binds to the A(+)/A (−) interface, though other antagonists may bind multiple interfaces. This suggests 5-$HT_{3A/B}$Rs have at least one A(+)/A(−) interface. However, atomic force microscopy experiments showed exogenously expressed receptors do not contain an A(+)/A(−) interface, but these experiments may not be physiologically representative (Barrera et al. 2005). 5-$HT_{3A/B}$Rs have a Hill coefficient close to one, which may suggest there is only one A(+)/A(−) interface. It is also possible that multiple stoichiometries are expressed similar to nAChR α2β4 receptors.

5-HT_3Rs are less sensitive to 5-HT than 5-HTRs in the GPCR subfamily. This allows for a differential physiological response where short-term spikes in 5-HT produce rapid transient signaling by 5-HT_3R and baseline 5-HT levels regulate slower metabotropic processes via other 5-HTRs. 5-HT_3Rs do have a lower threshold for 5-HT than other pLGICs have for their primary agonist. This is consistent with a modulatory role of 5-HT_3R as activation can occur at concentrations lower than those encountered during synaptic events. 5-HT_{3A}R also have a lower 5-HT activation threshold than 5-$HT_{3A/B}$R, consistent with 5-HT_{3A}R being more suited for presynaptic modulation than 5-$HT_{3A/B}$R (Dubin et al. 1999). 5-HT is the only endogenous full agonist of 5-HT_3R but there are other natural and synthetic ligands that act as agonists or partial agonists.

5-HT_3R antagonists generally bind with higher affinity than 5-HT. Setrons are an important class of 5-HT_3R antagonists. They are used clinically to treat chemotherapy or post-operative induced nausea and vomiting. This nausea and vomiting occurs in an acute and delayed phase. First generation setrons only treat acute nausea. These include alosetron, granisetron, ondansetron and tropisetron. The only second generation setron, palonosetron, treats both acute and delayed nausea (Rojas and Slusher 2012). Palonosetron is a competitive inhibitor like other setrons but is more effective because it remains bound to 5-HT_3R for up to several days, which has lasting effects on NK1 signaling, not entirely understood (Hothersall et al. 2013; Lummis and Thompson 2013; Wong et al. 1995; Hesketh et al. 2003). Setrons have also been investigated for the treatment of IBS though side-effects include severe constipation (Zheng et al. 2017). Alosetron is the only FDA-approved setron for IBS treatment. Other orthosteric antagonists act non-specifically on 5-HT_3R in addition to their primary targets. For example, the constipating effects of drugs such as morphine, methadone and anti-malarial drugs are at least partially mediated by 5-HT_3R (Baptista-Hon et al. 2012; Deeb et al. 2009; Thompson and Lummis 2008).

There are also several 5-HT_3R partial agonists that compete with 5-HT for binding, but elicit smaller maximal currents than 5-HT. Dopamine is an endogenous partial agonist of 5-HT_{3A}R that produces substantially less current than serotonin and favors

receptor deactivation to desensitization (Solt et al. 2007). This could lead to dampened but prolonged 5-HT_3R signaling at dopaminergic synapses. Several therapeutics also act as partial agonists. The smoking cessation agent Varenicline is a strong partial agonist of 5-HT_3R, which is thought to cause nausea (Lummis et al. 2011). While this is undesirable, weaker partial agonists are being investigated as useful therapeutics because they can reduce 5-HT_3R signaling without completely eliminating it (Roberts et al. 2020). This may prevent side effects sometimes caused by 5-HT_3R antagonists, such as severe constipation and ischemic colitis.

Mechanisms Underlying Ligand Binding

The key functional properties of 5-HT that allow it bind 5-HT_3R are the basic ethylamine group and the aromatic indole group. 5-HT_3R orthosteric antagonists have a similar structure to 5-HT, but generally have a bulkier ring-embedded basic nitrogen, sometimes a larger hydrophobic moiety, and extended linker between the two containing a hydrogen bond acceptor (Thompson 2013). The increased size and bonding partners give 5-HT_3R antagonists low nanomolar affinities compared to the micromolar affinity of 5-HT.

The free energy associated with ligand binding arises from both entropic and enthalpic contributions. Enthalpic contributions come from local interactions, such as hydrogen bonding, charge delocalization like pi-pi or pi-cation interactions, or electrostatic interactions. Entropic contributions come from increasing the conformational variability of the protein and surroundings. Though it is easy to understand how ligand binding decreases enthalpy, it may seem that ligand binding would also decrease entropy by restricting spatial flexibility and conformational variability. However, these entropy losses can be more than offset by solvent entropy, an essential consideration in many protein–ligand interactions. Changes in solvent-accessible surface area are a useful surrogate for changes in solvent entropy (Hilser et al. 2006). In particular, burying hydrophobic residues increases solvent entropy. Though empirical analysis of thermodynamic properties is beyond the scope of this chapter, it is useful to consider these principles in analyzing 5-HT_3R binding interactions. Note that all structures to date are from mouse 5-HT_{3A}R, though many of the interactions discussed are conserved in human 5-HT_{3A}R.

Entropic Contributions to Ligand Binding

The 5-HT_3R binding pocket is formed by the solvent-accessible gap between the protruding Loop C, Loop B and the (−) interface as shown in Fig. 11.5 (Basak et al. 2018b). This gap is lined with hydrophobic residues that give an entropic penalty to the apo conformation. Several of these residues are bulky aromatics that form the so-called aromatic box common to all pLGICs. These include F199 and Y207 from Loop C, W156 from Loop B, W63 from Loop D and Y126 from Loop E. The

shape of the aromatic box plays a role in neurotransmitter selectivity. For example, in GlyR and $GABA_AR$, the residues corresponding to W63 and W156 are either phenylalanine or tyrosine. This results in a smaller binding pocket more adapt to binding glycine or GABA (Du et al. 2015; Miller and Aricescu 2014). However, tryptophans are conserved in nAChR, consistent with acetylcholine's larger size (Morales-Perez et al. 2016).

In addition to the aromatic box, other hydrophobic residues also make entropic contributions to binding. They do not directly contact the ligand but their solvent-accessible area does change upon ligand binding. For example, conformational changes in Loop F and Loop C bury I180 against I201. Though neither residue directly contacts the ligand, their effect on binding is apparent in mutational studies (Zhang et al. 2007).

Comparing 5-HT-bound $5\text{-}HT_{3A}R$ structures with the granisetron-bound $5\text{-}HT_{3A}R$ structure further demonstrates the importance of solvent-accessible surface area (Basak et al. 2019). The 5-HT-bound structure only partially fills the binding pocket, allowing for conformational flexibility as shown in Fig. 11.5. The nitrogen-containing tropane moiety of granisetron more fully occupies this space. Additionally, the aromatic indazole group, similar in shape to serotonin, makes more contacts with the (−) interface due to the linker between the tropane and aromatic moieties. The large surface area of $5\text{-}HT_3R$ antagonists bury large portions of otherwise solvent-accessible surface area, which largely accounts for increased affinity of granisetron compared to 5-HT.

Enthalpic Contributions to Ligand Binding

$5\text{-}HT_3R$ mutagenesis studies have produced a long list of charged or hydrogen-bonding residues implicated in ligand binding. However, many of these residues are too spatially dispersed to simultaneously interact with the ligand. One possible explanation is that these spatially dispersed residues directly interact with bound ligand at different times in different conformations (Corrie and Baenziger 2013; Solt et al. 2007; Corradi et al. 2009). Another non-exclusive possibility is that water molecules coordinate indirect ligand–protein interactions within the binding pocket. Here we describe residues that structural or mutagenesis studies suggest make enthalpic contributions to the $5\text{-}HT_3R$ ligand binding, though the exact nature of these interactions is as yet undetermined. The position of these residues in 5-HT-bound or granisetron-bound structures are shown in Fig. 11.5.

Most enthalpic 5-HT interactions occur near the negatively charged pocket between Loop B and Loop C, which is conserved across pLGICs (Du et al. 2015; Laverty et al. 2019; Morales-Perez et al. 2016). In mouse $5\text{-}HT_{3A}R$, this pocket is formed by side chain oxygen atoms from N101, T154, Y207 and E209 and the main chain carbonyl oxygen from S155. Multiple 5-HT-bound structures show the ethylamine of 5-HT in a position near the side chain of T154 and the carbonyl oxygen of S155 (Basak et al. 2018a; Polovinkin et al. 2018).

Directly below this pocket are three Loop A residues, N101, E102 and F103 (Price et al. 2008; Sullivan et al. 2006; Boess et al. 1997; Steward et al. 2000; Nyce et al. 2010). In published structures, N101 faces towards the 5-HT ethylamine, but is too far away to directly interact. However, mutations to N101 significantly affect the EC_{50} of 5-HT (Nyce et al. 2010; Sullivan et al. 2006). It is possible that N101 interacts with 5-HT's ethylamine in a different conformation than that observed in present 5-HT-bound 5-HT_3R structures. This is observed in the 5-HT-bound structure of the soluble ECD surrogate of 5-HT_3R, 5-HTBP (Kesters et al. 2013). 5-HTBP is AChBP with an engineered agonist-binding pocket to mimic 5-HT_{3A}R. 5-HTBP has a tyrosine in the position of N101, but the tyrosine hydroxyl is positioned near the ethylamine of bound 5-HT. However, a 5-HT_{3A} N101V mutation improves the EC_{50} of 5-HT_{3A}R, inconsistent with possible hydrogen bonding (Price et al. 2008). E102 and F103 are positioned further away from the binding pocket than N101, but also have significant effects on 5-HT_3R binding (Sullivan et al. 2006; Boess et al. 1997). One intriguing possibility is that these form a less stable but more accessible 5-HT binding site that precedes a more stable binding conformation (Solt et al. 2007; Corradi et al. 2009). It is also possible that these residues play a role in coupling binding to channel gating.

Loop B lies above Loop A and forms part of the interior wall of the binding pocket. 5-HT induced currents are extremely sensitive to mutations from S150 through N164, but 5-HT_3R structures suggest only T154, S155 and W156 are exposed near the binding pocket (Basak et al. 2018a; Polovinkin et al. 2018). The other residues are likely involved with allosteric coupling or shaping the binding pocket. As previously mentioned, the side chain of T154 and carbonyl oxygen of S155 are positioned to form hydrogen bonds with the ethylamine of 5-HT in published 5-HT-bound 5-HT_{3A}R structures. Mutagenesis has also shown that W156 is critical for 5-HT binding (Spier and Lummis 2000; Beene et al. 2002; Thompson et al. 2008). This aromatic residue helps form the binding pocket as part of the aromatic box, but unnatural amino-acid mutagenesis shows that W156 also participates in pi-cation interactions with the 5-HT ethylamine. This suggests that the ethylamine can be in yet another conformation different from the 5-HT-bound 5-HT_3R structures, this one placing the ethylamine further out of the binding pocket from T154 and S155 towards the (−) interface.

The indole group of 5-HT is mostly involved in hydrophobic interactions in the aromatic box as previously discussed. However, the aromatic residues may also form pi-pi interactions with the indole rings. Additionally, the 5′ hydroxyl is positioned near the carbonyl oxygen of K127 and computational docking studies suggest several other 5′ hydroxyl positions that could contribute to ligand binding (Reeves et al. 2003).

Similar to entropic contributions, changes in protein conformation upon ligand binding also introduce enthalpic protein–protein interactions that make binding more favorable. Many of these interactions occur in residues across the binding interface. Y114 on Loop E is positioned to interact with H158 on Loop B upon ligand binding. Similarly, Y116 and Y126 on Loop E are positioned to hydrogen bond with Y207 on loop C in 5-HT-bound structures but not in apo or granisetron-bound structures. Unnatural mutagenesis shows that hydrogen bonding from both Y116 and Y126 is

important for channel function. It is likely that Loop C movements are coupled to Loop E via these hydrogen bonds (Beene et al. 2004; Venkataraman et al. 2002; Price and Lummis 2004). 5-HT_{3A}R structures show another possible cross-interface interaction between R65 on Loop D and D202 on Loop C. R65 is also positioned to bond with D42 on the β1 strand and R169 on Loop F. These residues are generally conserved across pLGICs and their mutation affects ligand binding (Yan et al. 1999; Thompson et al. 2006a).

The structurally observed enthalpic interactions between granisetron and 5-HT_3R are similar (Basak et al. 2019). The tropane nitrogen is in the ethylamine pocket. The carbonyl oxygen on the linker between tropane and the indazole moieties occupies almost the exact same position as the 5′ hydroxyl in the 5-HT-bound structure. This reinforces the indazole protruding towards the (−) interface more than the indole group of 5-HT as previously discussed. The protrusion changes interactions across the subunit interface compared to 5-HT binding and this may account for different effects on channel gating.

Conformational Changes Due to Ligand Binding

The binding of 5-HT or other ligands induces changes important for allosteric coupling, which is discussed in the next section. Here we address changes in the immediate vicinity of the binding pocket. The most obvious changes in protein structure come from Loop C movements. It has been shown that agonist binding in pLGICs induces an inward motion of Loop C relative to the apo state and most antagonists induce outward motion. However, it is unclear if these movements are necessary for channel gating, as discussed further in the section on allosteric coupling. The inward motion closes the gap between the two subunits and results in a smaller pocket accommodating the hydrophobic and hydrogen bond interactions previously discussed. Dissimilar to most pLGIC antagonists, the granisetron structure shows Loop C in-between the 5-HT-bound and apo positions (Basak et al. 2019). Perhaps setron inhibition of 5-HT_3R occurs via a mechanism distinct from that of other pLGIC antagonists.

Rearrangements in loop B are more subtle, but channel activity is very sensitive to mutations in this region. One source of protein rearrangements is the rotameric flexibility of W156. Both the 5-HT-bound 5-HTBP structure and the nanobody-stabilized 5-HT_3R crystal structure show W156 in a different conformation than later cryo-EM 5-HT_3R structures (Hassaine et al. 2014; Kesters et al. 2013). This may be an experimental artifact, but molecular dynamic simulations have been used to propose that this rotameric flexibility is important for coupling ligand binding and channel opening (Yuan et al. 2016).

Ligand binding also results in rearrangements in the relative position of neighboring subunits. When TMD helices are aligned, beta sheets 1, 2 (Loop D) and 6 (Loop E) are twisted counter-clockwise in granisetron-bound structures and more so in 5-HT-bound structures relative to the apo structure. This closes the gap between the interfaces around the bound ligand. The previously discussed interactions between

Loops B and C on the (+) interface and D, E and F on the (−) interface likely help mediate this beta-sheet twist.

There have been significant efforts to understand how changes in the ligand binding pocket are connected to channel activity. This is addressed more fully in the next section but here we note that each loop is covalently linked to loops in the ECD-TMD interface known to affect allosteric coupling. It seems likely there is not a single or even dominant conformational wave that connects ligand binding to channel activity, but rather there are rather multiple conformational pathways whose combinatorial effects determine the downstream effects of a specific ligand on channel activity.

Allosteric Coupling of Ligand Binding and Channel Gating

pLGICs are able to function as signal transducers because ligand binding is allosterically coupled to channel activity. The simple framework of pLGIC coupling is that channels go from an apo (resting) state, to a 5-HT-bound (open) state to a 5-HT-bound (desensitized) state before returning to the apo (closed) state. However, single-channel kinetic studies of open and closed dwell times reveal that this is an oversimplification and there are different lifetimes within each of these conformations. The lifetimes of these conformations determine the channel's overall kinetic properties. Generally, state lifetimes are measured by single-channel recordings, but WT 5-HT_{3A}R currents are too weak for these recordings. As an alternative, 5-HT_3R kinetic properties have been determined from macroscopic currents, albeit with less detail than single-channel recordings (Mott et al. 2001; Solt et al. 2007). Single-channel recordings have also been done with chimeric channels containing portions of 5-HT_{3A} or a high-conductance mutant of 5-HT_{3A}R that replaces the conductance-limiting intracellular arginines with their 5-HT_{3B} counterparts (Corradi et al. 2009; Bouzat et al. 2008). Both macroscopic current measurements and mutant recordings have limitations, but these studies provide valuable insight to 5-HT_3R kinetics and their physiological implications.

The mechanisms underlying coupling between ligand binding and channel activity are also not fully understood, but there are residues and protein structures known to be key contributors. These processes can be significantly altered by allosteric modulators. These modulators can enhance or inhibit channel function by stabilizing particular kinetic states. Allosteric modulators of 5-HT_3R have significant physiological consequences and include endogenous molecules, recreational drugs and potential therapeutics.

Observed Properties of Allosteric Coupling

5-HT_{3A}R has five identical binding sites and most evidence suggests that channel activation requires binding of at least two but more likely three molecules (Rayes et al. 2009; Mott et al. 2001; Corradi et al. 2009; Solt et al. 2007). Upon binding, the channel passes through a concentration-independent pre-open state. This occurs in multiple pLGICs, but transitions from the pre-open state to an open state are longer with 5-HT_3Rs than other channels. This accounts for their unusually long activation times, about 3 ms (Solt et al. 2007).

Electrophysiology recordings showed there are multiple open-state lifetimes and it was originally proposed that different lifetimes corresponded to different binding occupancies (Mott et al. 2001). Three open-state lifetimes were observed in the high-conductance 5-HT_{3A}R mutant but their relative ratios did not change with 5-HT concentration, suggesting all three states are accessible regardless of occupancy (Corradi et al. 2009). As an exception, the duration of the longest open-state lifetime is shortened at higher 5-HT concentrations (>5 μM), which is consistent with 5-HT pore-block.

Single channel recordings of chimeric or mutant 5-HT_3R receptors show bursts of openings separated by short channel closures, typical of most ion channels (Sakmann et al. 1980). A less common feature of 5-HT_{3A}R is that multiple bursts are separated by intermediate closures that are known as clusters. The cluster terminates with a long closure before activity resumes (Bouzat et al. 2008; Corradi et al. 2009). The intermediate closures are thought to be transitions back to the pre-open state and the long closure is a desensitized state. The effect of clustering prolongs channel activity. Indeed, macroscopic-current studies of WT 5-HT_3R show that the channel desensitizes with a time constant of 1 s, which is slower than most pLGICs. Complete desensitization recovery occurs in sigmoidal fashion over about 20 s with a Hill coefficient of about 3. This is consistent with the 5-HT_{3A}R sequentially unbinding 3 5-HT molecules before returning to a closed state that can be reactivated. The rate of desensitization also increases as a function of 5-HT and extracellular calcium, perhaps through mechanisms similar to 5-HT pore block and inhibition by extracellular calcium (Yakel et al. 1993; Solt et al. 2007). The kinetics of 5-HT_{3A}R are too slow for postsynaptic transmission, but are consistent with a mostly presynaptic modulatory role of 5-HT_{3A}Rs that persists through several synaptic events.

5-$HT_{3A/B}$Rs have not been examined to the same degree as 5-HT_{3A}R, but they have roughly a three times lower Hill coefficient for 5-HT activation, suggesting activation requires fewer bound 5-HT molecules. Heteromeric receptors also desensitize and recover more quickly than 5-HT_{3A}Rs, consistent with 5-$HT_{3A/B}$R being better suited for postsynaptic transmission (Hapfelmeier et al. 2003; Dubin et al. 1999).

Mechanisms Underlying Coupling Between Ligand Binding and Ion Permeation

For all pLGICs, loop regions in the ligand-binding domain are covalently connected to four ECD regions at the ECD-TMD interface that are critical for allosteric coupling, the β1-β2 loop, the β6-β7 (cys) loop, β6-β7 loop and the pre-M1 helix. Movements in these regions are communicated to the TMD via covalent M1 linkage and interactions with the M2-M3 linker (Bouzat et al. 2004). Thermodynamic analysis of mutant perturbations shows that a 'conformational wave' proceeds from the ligand binding pocket, to ECD portions of the ECD-TMD regions, to the M2-M3 linker and eventually the pore (Chakrapani et al. 2004). Beyond this crude outline, the events communicating ligand binding to the pore domain remain unclear.

It was originally proposed that movements in the mobile Loop C displaced the pre-M1 helix which in turn moved the β1-β2 and β6-β7 loops that finally repositioned the M2-M3 linker (Bouzat et al. 2004; Lee et al. 2009). However, there is uncertainty over the role of Loop C in channel opening with reports differing among pLGICs (Terejko et al. 2020; Purohit and Auerbach 2013; Pless and Lynch 2009). It is possible that conformational changes in the ligand binding pocket act on multiple kinetic pathways whose combinatorial effects determine the channel state, like the strings of a puppet. To support this idea, 5-HT_3R antagonist-bound structures resemble open state structures in the peripheral beta sheets, including Loop C and the pre-M1 helix, and resemble closed state structures in the inner beta sheets, including the β1-β2 loop as seen in Fig. 11.6.

Exact residues at the ECD-TMD interface are poorly conserved across pLGICs, but general principles are observed (Mukhtasimova and Sine 2013; Xiu et al. 2005). The charged residues at the ECD-TMD interface form a network of interactions with net charge varying between structures. The spatial layout of 5-HT_3R structures suggest that the pre-M1 helix contacts the β1-β2, β6-β7 and β8-β9 loops, and that each of these loops contacts the M2-M3 linker, where movements are directly coupled to the channel pore. The pre-M1 and M1 helices are also positioned to directly interact with the M2-M3 linker.

The pre-M1 helix carries a net positive charge including three consecutive arginine residues R217-219, of which R218 is conserved across pLGICs. Structural evidence places these residues at a hub between E53 on the β1-β2 loop, D145 on the β6-β7 loop and E186 and W187 on the β8-β9 loop (Hassaine et al. 2014; Basak et al. 2018b). Mutation of these residues affects channel gating or formation. For mutations that disrupt channel formation, similar mutations in other pLGICs affect gating, though the effects vary from one member to the next (Price et al. 2007; Bouzat et al. 2004; Hu et al. 2003).

The E53 connection between the pre-M1 helix and β1-β2 loop creates a possible conformational pathway along the inner vestibule. Mutational experiments provide evidence that this pathway continues to the M2-M3 linker via K54. K54A mutants bind ligand but do not gate while K54R or K54Q mutations retain gating (Mosesso

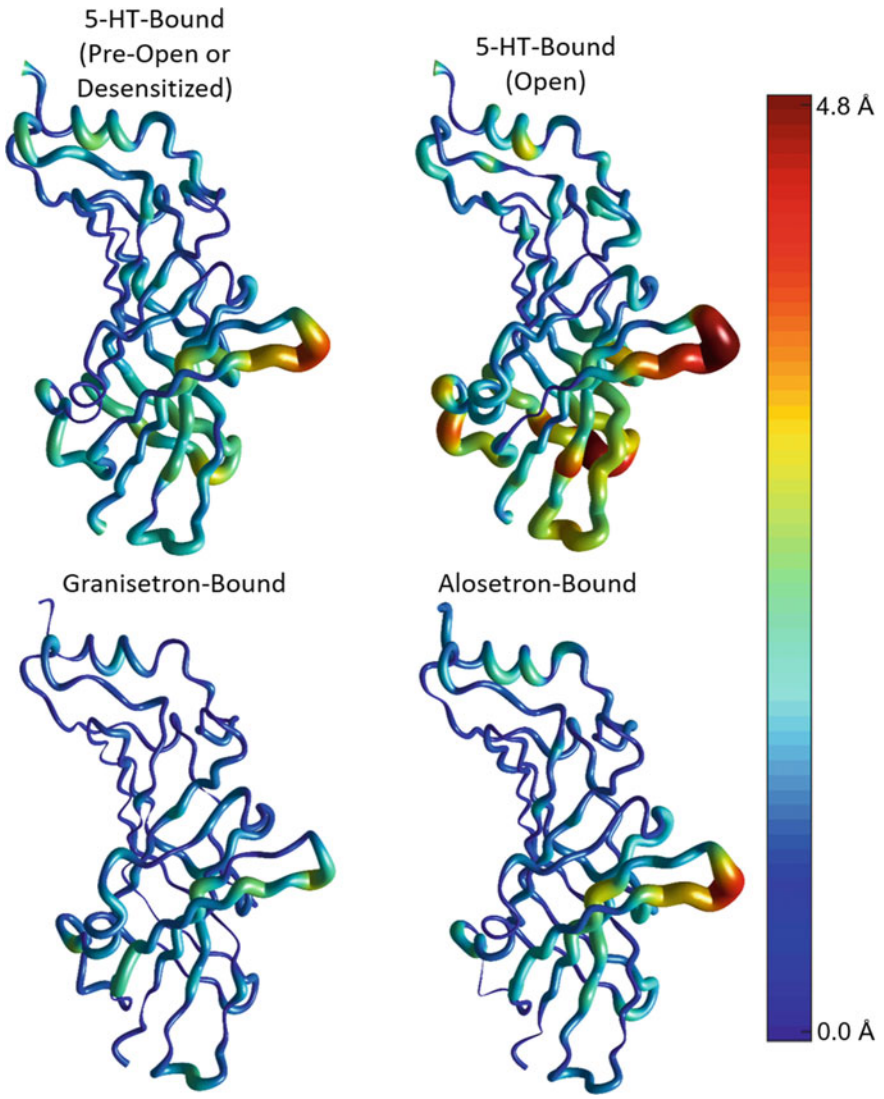


Fig. 11.6 Allosterically coupled motions of the ECD domain upon ligand binding. The ECD of each labeled structure was aligned to the ECD of apo 5-HT_{3A}R (PDBs 6BE1, 6DG7, 6DG8, 6NP0 and 6W1J) (Basak et al. 2018a, b, 2019, 2020). The rendered images show the Cα chain with a thickness and color that correspond to the relative displacements from the apo structure. Note that some regions show similar displacements in 5-HT and antagonist bound structures, such as Loop C, while other motions are unique to 5-HT bound structures, such as the β1-β2 loop.

et al. 2019). K54C mutations have disrupted gating that can be rescued by sulfhydryl-labelling agents, regardless of charge (Reeves et al. 2005). Additionally, double mutants with A277C or I278C, near M2 on the M2-M3 linker, are completely inhibited, but currents are rescued in the presence of reducing agents. Together this suggests that K54 sterically interacts with residues near the top of the M2 helix.

The β6-β7 loop provides a connection to the M2-M3 linker along the outer vestibule. This loop consists of several hydrophobic residues including the highly conserved F142, P143 and F144, or FPF motif. Structural and mutagenesis experiments suggest that these residues penetrate into the membrane interface and push L282 and I283 upon channel activation (Hassaine et al. 2014; Limapichat et al. 2010; Lee et al. 2009). These residues are located at the M3 end of the M2-M3 linker next to P281. P281 is conserved across cationic pLGICs and plays a key role in gating. Specifically, natural and unnatural mutagenesis show that propensity of the residue to be in the cis conformation correlates with gating efficiency, though isomerization is not required for channel opening (Paulsen et al. 2009; Lummis et al. 2005; Mosesso et al. 2019). The mobility of the β1-β2 and β6-β7 loops in 5-HT-bound states compared to antagonist-bound states can be observed in Fig. 11.6.

The M2-M3 linker is displaced radially outwards and towards the membrane in the in open 5-HT_3 structures. This is accompanied by a rotation and tilting of the M2 helix that widens the pore at the hydrophobic gate. In addition to the effects of the M2-M3 linker, there are several mutations throughout the TMD helices and ICD that affect channel gating, especially desensitization (Gunthorpe et al. 2000; Yakel et al. 1993). It has been proposed that movements in the M1 or M3 helix are communicated to M2 residues via charged residue pairs within the helix bundle. The length and composition of the M3-M4 linker also affect desensitization, perhaps through mechanisms that alter conduction through the intracellular portals (McKinnon et al. 2012; Hu et al. 2006).

Allosteric Modulators

Allosteric modulators bind to 5-HT_3Rs non-competitively and stabilize kinetic states that can either enhance or reduce channel properties. Positive allosteric modulators (PAMs) will not directly activate the channel but will potentiate the receptor so it activates at lower agonist concentrations. This has significant physiological consequences. For example, small alcohols and halogenated anesthetics are PAMs of several pLGICs including $GABA_A$Rs and 5-HT_{3A}Rs (Jenkins et al. 1996; Lovinger 1991). At GABAergic synapses in inhibitory neurons, presynaptic activation of potentiated 5-HT_{3A}Rs results in increased GABA release, which activates potentiated postsynaptic $GABA_A$Rs. Similarly, 5-HT_{3A}R potentiation increases dopamine release at dopaminergic synapses (Wozniak et al. 1990). These neuronal events underlie the sensations caused by these compounds. Defects in 5-HT_3R signaling are connected to various psychiatric disorders and substance abuse and 5-HT_3R

PAMs are attractive drug candidates to address these conditions (Thompson et al. 2006b; Hagan et al. 1993).

The binding cavity for alcohols and volatile anesthetics is located between M3 and M1 helices in neighboring subunits at the ECD-TMD interface. Binding at this site stabilizes the channel open state (Feinberg-Zadek and Davies 2010; Li et al. 2006). At higher concentrations, these compounds bind a distinct low-affinity site and act as negative allosteric modulators (NAMs) (Rüsch et al. 2007). This second site may act as a physiological override to excessive drug concentrations. 5-$HT_{3A/B}$R is also potentiated by alcohol and anesthetics but to a lesser degree, consistent with presynaptic 5-HT_3Rs being 5-HT_{3A}Rs (Solt et al. 2005).

There are many other natural 5-HT_3R NAMs. These include cannabinoids, menthol and other terpenes, ginger extracts and capsaicin. Both endogenous cannabinoids such as arachidonoyl glycerol and arachidonoyl ethanolamine (anandamide) and exogenous cannabinoids such as cannabidiol (CBD) and Δ^9-tetrahydrocannabinol (THC) inhibit 5-HT_3R currents (Yang et al. 2010; Barann et al. 2002; Xiong et al. 2008). Notably, CBD effectively inhibits 5-HT_3R but has no psychoactive effects. Cannabinoids interact directly with 5-HT_3R and inhibit channel function by enhanced desensitization. The mechanism of inhibition is not fully understood but in GlyR, cannabinoids bind S296 within the M3 helix (Xiong et al. 2011). 5-HT_3R activity mediates analgesic and anti-emetic effects of cannabinoids in CB_1 and CB_2 KO mice (Rácz et al. 2008). This suggests that further developed 5-HT_3R modulators may offer the therapeutic benefits of cannabinoids without psychoactive effects.

The effects of alcohols and cannabinoids are applicable to many natural compounds found in traditional treatments for nausea and pain. Small terpenes and terpenoids have chemical structures similar to larger alcohols or anesthetics. Consistent with their size, molecules like menthol, boldine, citronellol and geraniol are NAMs with a range of potencies (Walstab et al. 2014; Ziemba et al. 2015). However, other similarly size molecules such as thymol and carvacrol are PAMs. Other natural NAMs include ginger and capsaicin. These lipid-embedded molecules may act through similar mechanisms as cannabinoids. These natural compounds act on many molecular targets, but 5-HT_3R modulation may contribute to their physiological effects. The range of potency and directional effects suggests that further study of these compounds will lead to a better understanding of allosteric mechanisms and perhaps useful pharmacological tools.

Conclusions

It has been about thirty years since the discovery 5-HT_{3A}R as an ion channel (Derkach et al. 1989; Maricq et al. 1991). At that time, questions were focused on the nature of 5-HT_3Rs. Where are they located? What do they look like? What are they basic properties? As covered in this chapter, substantial progress has been made on many of these questions. Still, several fundamental properties of 5-HT_3R remain unknown.

For example, most the ICD remains a black box even in recent cryo-EM structures. This domain is the most divergent region between pLGICs and is the site of many protein–protein interactions that regulate channel function, expression and localization. Future research is required to better understand this and other fundamental parts of 5-HT_3R.

Our increased understanding of what the 5-HT_3R is and does has led to further questions of how the channel operates. For example: Why is calcium permeable in 5-HT_{3A}R but not 5-$HT_{3A/B}$R? Why do some ligands induce channel opening while others antagonize it? What are the events that lead from ligand binding to channel opening? Answers to these questions are more than just intellectual curiosities, they help us understand channel function or dysfunction in certain conditions and provide a template for pharmacological intervention.

Ion channels seldom operate alone. The function of homomeric or heteromeric channel assembly is largely influenced by local environments that include membrane lipids, interacting protein partners related to the extracellular matrix, cytoskeletal proteins and cellular signaling molecules. While ion channels are now studied at the structural single-molecule level, future directions will increasingly be focused on the structure and function of interactome assemblies, and ultimately proteins in their native cellular environments.

A better understanding of how the channel operates will hopefully lead to an increased ability to pharmacologically manipulate its function. There is a growing list of natural and synthesized compounds that affect 5-HT_3R. In particular, there is promise in the use of weak partial agonists and NAMs as anti-nausea medications. An increased library of specific PAMs would be useful for understanding the role 5-HT_3R in synaptic signaling as present PAMs are non-specific.

Even an exact understanding of 5-HT_3R itself would be of limited practical use, since the channel is clearly tailored to produce specific events during signaling, but the downstream effects are at best crudely understood. Future research on the implications of 5-HT_3R expression and activity throughout the body can open the door to understanding complex neurobehaviors and addressing devastating mental disease. This is an immense challenge but the potential benefits touch on the ultimate aim of biomedical research.

Acknowledgements The authors would like to acknowledge support from members of the Chakrapani lab in discussing and editing the publication as well as Whitney Gibbs for the figure illustrations.

References

Babic T, Troy AE, Fortna SR, Browning KN (2012) Glucose-dependent trafficking of 5-HT3 receptors in rat gastrointestinal vagal afferent neurons. Neurogastroenterol Motil 24(10):e476–e488

Babic T, Browning KN (2014) The role of vagal neurocircuits in the regulation of nausea and vomiting. Eur J Pharmacol 722:38–47

Baptista V, Browning KN, Travagli RA (2007) Effects of cholecystokinin-8s in the nucleus tractus solitarius of vagally deafferented rats. Am J Phys-Regul, Integr Comp Physiol 292(3):R1092–R1100

Baptista-Hon DT, Deeb TZ, Othman NA, Sharp D, Hales TG (2012) The 5-HT3B subunit affects high-potency inhibition of 5-HT3 receptors by morphine. Br J Pharmacol 165(3):693–704

Barann M, Molderings G, Brüss M, Bönisch H, Urban B, Göthert M (2002) Direct inhibition by cannabinoids of human 5-HT3A receptors: probable involvement of an allosteric modulatory site. Br J Pharmacol 137(5):589–596

Barrera NP, Herbert P, Henderson RM, Martin IL, Edwardson JM (2005) Atomic force microscopy reveals the stoichiometry and subunit arrangement of 5-HT3 receptors. Proc Natl Acad Sci 102(35):12595–12600

Basak S, Gicheru Y, Rao S, Sansom MS, Chakrapani S (2018a) Cryo-EM reveals two distinct serotonin-bound conformations of full-length 5-HT 3A receptor. Nature 563(7730):270

Basak S, Gicheru Y, Samanta A, Molugu SK, Huang W, la de Fuente M, Hughes T, Taylor DJ, Nieman MT, Moiseenkova-Bell V (2018b) Cryo-EM structure of 5-HT 3A receptor in its resting conformation. Nat Commun 9(1):514

Basak S, Gicheru Y, Kapoor A, Mayer ML, Filizola M, Chakrapani S (2019) Molecular mechanism of setron-mediated inhibition of full-length 5-HT 3A receptor. Nat Commun 10(1):1–11

Basak S, Kumar A, Ramsey S, Gibbs E, Kapoor A, Filizola M, Chakrapani S (2020) High-resolution Structures of multiple 5-HT3AR-setron complexes reveal a novel mechanism of competitive inhibition. eLife 9:e57870. https://doi.org/10.7554/eLife.57870

Beene DL, Brandt GS, Zhong W, Zacharias NM, Lester HA, Dougherty DA (2002) Cation−π interactions in ligand recognition by serotonergic (5-HT3A) and nicotinic acetylcholine receptors: the anomalous binding properties of nicotine. Biochemistry 41(32):10262–10269

Beene DL, Price KL, Lester HA, Dougherty DA, Lummis SC (2004) Tyrosine residues that control binding and gating in the 5-hydroxytryptamine-3 receptor revealed by unnatural amino acid mutagenesis. J Neurosci 24(41):9097–9104

Bertrand P, Kunze W, Furness J, Bornstein J (2000) The terminals of myenteric intrinsic primary afferent neurons of the guinea-pig ileum are excited by 5-hydroxytryptamine acting at 5-hydroxytryptamine-3 receptors. Neuroscience 101(2):459–469

Blackshaw L, Grundy D (1993) Effects of 5-hydroxytryptamine (5-HT) on the discharge of vagal mechanoreceptors and motility in the upper gastrointestinal tract of the ferret. J Auton Nerv Syst 45(1):51–59

Boess F, Steward L, Steele J, Liu D, Reid J, Glencorse T, Martin I (1997) Analysis of the ligand binding site of the 5-HT3 receptor using site directed mutagenesis: importance of glutamate 106. Neuropharmacology 36(4–5):637–647

Bouzat C, Gumilar F, Spitzmaul G, Wang H-L, Rayes D, Hansen SB, Taylor P, Sine SM (2004) Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel. Nature 430(7002):896–900

Bouzat C, Bartos M, Corradi J, Sine SM (2008) The interface between extracellular and transmembrane domains of homomeric Cys-loop receptors governs open-channel lifetime and rate of desensitization. J Neurosci 28(31):7808–7819

Boyd GW, Low P, Dunlop JI, Robertson LA, Vardy A, Lambert JJ, Peters JA, Connolly CN (2002) Assembly and cell surface expression of homomeric and heteromeric 5-HT3 receptors: the role of oligomerization and chaperone proteins. Mol Cell Neurosci 21(1):38–50

Brady CA, Dover TJ, Massoura AN, Princivalle AP, Hope AG, Barnes NM (2007) Identification of 5-HT3A and 5-HT3B receptor subunits in human hippocampus. Neuropharmacology 52(5):1284–1290

Brejc K, van Dijk WJ, Smit AB, Sixma TK (2002) The 2.7 Å Structure of AChBP, Homologue of the ligand-binding domain of the nicotinic acetylcholine receptor. In: Ion channels: from atomic resolution physiology to functional genomics: novartis foundation symposium 245. Wiley Online Library, pp 22–32

Brown A, Hope A, Lambert J, Peters J (1998) Ion permeation and conduction in a human recombinant 5-HT3 receptor subunit (h5-HT3A). J Phys 507(3):653–665

Brüss M, Barann M, Hayer-Zillgen M, Eucker T, Göthert M, Bönisch H (2000) Modified 5-HT 3A receptor function by co-expression of alternatively spliced human 5-HT 3A receptor isoforms. Naunyn-Schmiedeberg's Arch Pharmacol 362(4–5):392–401

Bufton KE, Steward LJ, Barber PC, Barnes NM (1993) Distribution and characterization of the [3H] granisetron-labelled 5-HT3 receptor in the human forebrain. Neuropharmacology 32(12):1325–1331

Carland JE, Cooper MA, Sugiharto S, Jeong H-J, Lewis TM, Barry PH, Peters JA, Lambert JJ, Moorhouse AJ (2009) Characterization of the effects of charged residues in the intracellular loop on ion permeation in α1 glycine receptor channels. J Biol Chem 284(4):2023–2030

Chakrapani S, Bailey TD, Auerbach A (2004) Gating dynamics of the acetylcholine receptor extracellular domain. J Gen Phys 123(4):341–356

Cheng A, Bollan KA, Greenwood SM, Irving AJ, Connolly CN (2007) Differential subcellular localization of RIC-3 isoforms and their role in determining 5-HT3 receptor composition. J Biol Chem 282(36):26158–26166

Choi IS, Cho JH, Kim JT, Park EJ, Lee MG, Shin HI, Choi BJ, Jang IS (2007) Serotoninergic modulation of GABAergic synaptic transmission in developing rat CA3 pyramidal neurons. J Neurochem 103(6):2342–2353

Corradi J, Gumilar F, Bouzat C (2009) Single-channel kinetic analysis for activation and desensitization of homomeric 5-HT3A receptors. Biophys J 97(5):1335–1345

Corrie J, Baenziger JE (2013) Gating of pentameric ligand-gated ion channels: structural insights and ambiguities. Structure 21(8):1271–1283

Corry B (2006) An energy-efficient gating mechanism in the acetylcholine receptor channel suggested by molecular and Brownian dynamics. Biophys J 90(3):799–810

Dang H, England PM, Farivar SS, Dougherty DA, Lester HA (2000) Probing the role of a conserved M1 proline residue in 5-hydroxytryptamine3 receptor gating. Mol Pharmacol 57(6):1114–1122

Das P, Dillon GH (2003) The 5-HT3B subunit confers reduced sensitivity to picrotoxin when co-expressed with the 5-HT3A receptor. Mol Brain Res 119(2):207–212

Das P, Dillon GH (2005) Molecular determinants of picrotoxin inhibition of 5-hydroxytryptamine type 3 receptors. J Pharmacol Exp Ther 314(1):320–328

Davies PA, Pistis M, Hanna MC, Peters JA, Lambert JJ, Hales TG, Kirkness EF (1999) The 5-HT 3B subunit is a major determinant of serotonin-receptor function. Nature 397(6717):359–363

Deeb TZ, Carland JE, Cooper MA, Livesey MR, Lambert JJ, Peters JA, Hales TG (2007) Dynamic modification of a mutant cytoplasmic cysteine residue modulates the conductance of the human 5-HT3A receptor. J Biol Chem 282(9):6172–6182

Deeb TZ, Sharp D, Hales TG (2009) Direct subunit-dependent multimodal 5-hydroxytryptamine3 receptor antagonism by methadone. Mol Pharmacol 75(4):908–917

Derkach V, Surprenant A, North R (1989) 5-HT3 receptors are membrane ion channels. Nature 339(6227):706–709

Dorostkar MM, Boehm S (2007) Opposite effects of presynaptic 5-HT3 receptor activation on spontaneous and action potential-evoked GABA release at hippocampal synapses. J Neurochem 100(2):395–405

Dubin AE, Huvar R, D'Andrea MR, Pyati J, Zhu JY, Joy K, Wilson SJ, Galindo JE, Glass CA, Luo L (1999) The pharmacological and functional characteristics of the serotonin 5-HT3A receptor are specifically modified by a 5-HT3B receptor subunit. J Biol Chem 274(43):30799–30810

Du J, Lü W, Wu S, Cheng Y, Gouaux E (2015) Glycine receptor mechanism elucidated by electron cryo-microscopy. Nature 526(7572):224

Emerit M, Martres M, Miquel M, El Mestikawy S, Hamon M (1995) Differentiation alters the expression of the two splice variants of the serotonin 5-HT3 receptor-A mRNA in NG108-15 Cells. J Neurochem 65(5):1917–1925

Everitt AB, Seymour VA, Curmi J, Laver DR, Gage PW, Tierney ML (2009) Protein interactions involving the γ2 large cytoplasmic loop of GABAA receptors modulate conductance. FASEB J 23(12):4361–4369

Fakhfouri G, Rahimian R, Dyhrfjeld-Johnsen J, Zirak MR, Beaulieu J-M (2019) 5-HT3 Receptor antagonists in neurologic and neuropsychiatric disorders: the Iceberg still lies beneath the surface. Pharmacol Rev 71(3):383–412

Feinberg-Zadek PL, Davies PA (2010) Ethanol stabilizes the open state of single 5-hydroxytryptamine3A (QDA) receptors. J Pharmacol Exp Ther 333(3):896–902

Férézou I, Cauli B, Hill EL, Rossier J, Hamel E, Lambolez B (2002) 5-HT3 receptors mediate serotonergic fast synaptic excitation of neocortical vasoactive intestinal peptide/cholecystokinin interneurons. J Neurosci 22(17):7389–7397

Filatov GN, White MM (1995) The role of conserved leucines in the M2 domain of the acetylcholine receptor in channel gating. Mol Pharmacol 48(3):379–384

Gill CH, Peters JA, Lambert JJ (1995) An electrophysiological investigation of the properties of a murine recombinant 5-HT3 receptor stably expressed in HEK 293 cells. Br J Pharmacol 114(6):1211–1221

Glatzle J, Sternini C, Robin C, Zittel TT, Wong H, Reeve JR Jr, Raybould HE (2002) Expression of 5-HT3 receptors in the rat gastrointestinal tract. Gastroenterology 123(1):217–226

Glaum SR, Proudfit HK, Anderson EG (1990) 5-HT3 receptors modulate spinal nociceptive reflexes. Brain Res 510(1):12–16

Glaum SR, Brooks PA, Spyer KM, Miller RJ (1992) 5-Hydroxytryptamine-3 receptors modulate synaptic activity in the rat nucleus tractus solitarius in vitro. Brain Res 589(1):62–68

Grabauskas G, Song I, Zhou S, Owyang C (2010) Electrophysiological identification of glucose-sensing neurons in rat nodose ganglia. J Physiol 588(4):617–632

Gumilar F, Bouzat C (2008) Tricyclic antidepressants inhibit homomeric cys-loop receptors by acting at different conformational states. Eur J Pharmacol 584(1):30–39

Gunthorpe MJ, Peters JA, Gill CH, Lambert JJ, Lummis SC (2000) The 4′ lysine in the putative channel lining domain affects desensitization but not the single-channel conductance of recombinant homomeric 5-HT3A receptors. J Physiol 522(2):187–198

Gunthorpe MJ, Lummis SC (2001) Conversion of the ion selectivity of the 5-HT3Areceptor from cationic to anionic reveals a conserved feature of the ligand-gated ion channel superfamily. J Biol Chem 276(14):10977–10983

Hagan R, Kilpatrick G, Tyers M (1993) Interactions between 5-HT 3 receptors and cerebral dopamine function: implications for the treatment of schizophrenia and psychoactive substance abuse. Psychopharmacology 112(1):S68–S75

Hales TG, Dunlop JI, Deeb TZ, Carland JE, Kelley SP, Lambert JJ, Peters JA (2006) Common determinants of single channel conductance within the large cytoplasmic loop of 5-hydroxytryptamine type 3 and α4β2 nicotinic acetylcholine receptors. J Biol Chem 281(12):8062–8071

Hansen SB, Wang H-L, Taylor P, Sine SM (2008) An ion selectivity filter in the extracellular domain of Cys-loop receptors reveals determinants for ion conductance. J Biol Chem 283(52):36066–36070

Hapfelmeier G, Tredt C, Haseneder R, Zieglgänsberger W, Eisensamer B, Rupprecht R, Rammes G (2003) Co-expression of the 5-HT3B serotonin receptor subunit alters the biophysics of the 5-HT3 receptor. Biophys J 84(3):1720–1733

Hassaine G, Deluz C, Grasso L, Wyss R, Tol MB, Hovius R, Graff A, Stahlberg H, Tomizaki T, Desmyter A (2014) X-ray structure of the mouse serotonin 5-HT 3 receptor. Nature 512(7514):276–281

Hénault CM, Sun J, Therien JD, Corrie J, Carswell CL, Labriola JM, Juranka PF, Baenziger JE (2015) The role of the M4 lipid-sensor in the folding, trafficking, and allosteric modulation of nicotinic acetylcholine receptors. Neuropharmacology 96:157–168

Hesketh P, Van Belle S, Aapro M, Tattersall F, Naylor R, Hargreaves R, Carides A, Evans J, Horgan K (2003) Differential involvement of neurotransmitters through the time course of cisplatin-induced emesis as revealed by therapy with specific receptor antagonists. Eur J Cancer 39(8):1074–1080

Hilf RJ, Dutzler R (2008) X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature 452(7185):375–379
Hillsley K, Grundy D (1998) Sensitivity to 5-hydroxytryptamine in different afferent subpopulations within mesenteric nerves supplying the rat jejunum. J Physiol 509(3):717–727
Hillsley K, Kirkup A, Grundy D (1998) Direct and indirect actions of 5-hydroxytryptamine on the discharge of mesenteric afferent fibres innervating the rat jejunum. J Physiol 506(2):551–561
Hilser VJ, García-Moreno EB, Oas TG, Kapp G, Whitten ST (2006) A statistical thermodynamic model of the protein ensemble. Chem Rev 106(5):1545–1558
Hothersall JD, Moffat C, Connolly CN (2013) Prolonged inhibition of 5-HT3 receptors by palonosetron results from surface receptor inhibition rather than inducing receptor internalization. Br J Pharmacol 169(6):1252–1262
Hu X-Q, Zhang L, Stewart RR, Weight FF (2003) Arginine 222 in the pre-transmembrane domain 1 of 5-HT3A receptors links agonist binding to channel gating. J Biol Chem 278(47):46583–46589
Hu XQ, Lovinger DM (2005) Role of aspartate 298 in mouse 5-HT3A receptor gating and modulation by extracellular Ca2+. J Physiol 568(2):381–396
Hu X-Q, Sun H, Peoples RW, Hong R, Zhang L (2006) An interaction involving an arginine residue in the cytoplasmic domain of the 5-HT3A receptor contributes to receptor desensitization mechanism. J Biol Chem 281(31):21781–21788
Hussy N, Lukas W, Jones KA (1994) Functional properties of a cloned 5-hydroxytryptamine ionotropic receptor subunit: comparison with native mouse receptors. J Physiol 481(2):311–323
Imoto K, Busch C, Sakmann B, Mishina M, Konno T, Nakai J, Bujo H, Mori Y, Fukuda K, Numa S (1988) Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. Nature 335(6191):645–648
Jeggo RD, Kellett DO, Wang Y, Ramage AG, Jordan D (2005) The role of central 5-HT3 receptors in vagal reflex inputs to neurones in the nucleus tractus solitarius of anaesthetized rats. J Physiol 566(3):939–953
Jenkins A, Franks N, Lieb W (1996) Actions of general anaesthetics on 5-HT3 receptors in N1E–115 neuroblastoma cells. Br J Pharmacol 117(7):1507–1515
Jensen AA, Davies PA, Bräuner-Osborne H, Krzywkowski K (2008) 3B but which 3B? And that's just one of the questions: the heterogeneity of human 5-HT3 receptors. Trends Pharmacol Sci 29(9):437–444
Jones B, Costall B, Domeney A, Kelly M, Naylor R, Oakley N, Tyers M (1988) The potential anxiolytic activity of GR38032F, a 5-HT3-receptor antagonist. Br J Pharmacol 93(4):985–993
Jones KA, Surprenant A (1994) Single channel properties of the 5-HT3 subtype of serotonin receptor in primary cultures of rodent hippocampus. Neurosci Lett 174(2):133–136
Kapeller J, Möller D, Lasitschka F, Autschbach F, Hovius R, Rappold G, Brüss M, Gershon MD, Niesler B (2011) Serotonin receptor diversity in the human colon: Expression of serotonin type 3 receptor subunits 5-HT3C, 5-HT3D, and 5-HT3E. J Comp Neurol 519(3):420–432
Keating DJ, Spencer NJ (2010) Release of 5-hydroxytryptamine from the mucosa is not required for the generation or propagation of colonic migrating motor complexes. Gastroenterology 138(2):659–670. e652
Kelley SP, Dunlop JI, Kirkness EF, Lambert JJ, Peters JA (2003) A cytoplasmic region determines single-channel conductance in 5-HT 3 receptors. Nature 424(6946):321–324
Kesters D, Thompson AJ, Brams M, Van Elk R, Spurny R, Geitmann M, Villalgordo JM, Guskov A, Danielson UH, Lummis SC (2013) Structural basis of ligand recognition in 5-HT3 receptors. EMBO Rep 14(1):49–56
Koyama S, Matsumoto N, Kubo C, Akaike N (2000) Presynaptic 5-HT3 receptor-mediated modulation of synaptic GABA release in the mechanically dissociated rat amygdala neurons. J Physiol 529(2):373–383
Lacolley P, Owen J, Sandock K, Lewis T, Bates J, Robertson T, Lewis S (2006a) 5-HT activates vagal afferent cell bodies in vivo: role of 5-HT2 and 5-HT3 receptors. Neuroscience 143(1):273–287

Lacolley P, Owen J, Sandock K, Lewis T, Bates J, Robertson T, Lewis S (2006b) Occipital artery injections of 5-HT may directly activate the cell bodies of vagal and glossopharyngeal afferent cell bodies in the rat. Neuroscience 143(1):289–308

Laverty D, Desai R, Uchański T, Masiulis S, Stec WJ, Malinauskas T, Zivanov J, Pardon E, Steyaert J, Miller KW (2019) Cryo-EM structure of the human α1β3γ2 GABA A receptor in a lipid bilayer. Nature 565(7740):516

Leal-Cardoso H, Koschorke GM, Taylor G, Weinreich D (1993) Electrophysiological properties and chemosensitivity of acutely isolated nodose ganglion neurons of the rabbit. J Auton Nerv Syst 45(1):29–39

Lee WY, Free CR, Sine SM (2009) Binding to gating transduction in nicotinic receptors: cys-loop energetically couples to pre-M1 and M2–M3 regions. J Neurosci 29(10):3189–3199

Liang D-Y, Li X, Clark JD (2011) 5-hydroxytryptamine type 3 receptor modulates opioid-induced hyperalgesia and tolerance in mice. Anesthesiology 114(5):1180

Limapichat W, Lester HA, Dougherty DA (2010) Chemical scale studies of the Phe-Pro conserved motif in the cys loop of Cys loop receptors. J Biol Chem 285(12):8976–8984

Livesey MR, Cooper MA, Deeb TZ, Carland JE, Kozuska J, Hales TG, Lambert JJ, Peters JA (2008) Structural determinants of Ca2+ permeability and conduction in the human 5-hydroxytryptamine type 3A receptor. J Biol Chem 283(28):19301–19313

Livesey MR, Cooper MA, Lambert JJ, Peters JA (2011) Rings of charge within the extracellular vestibule influence ion permeation of the 5-HT3A receptor. J Biol Chem 286(18):16008–16017

Li G-D, Chiara DC, Sawyer GW, Husain SS, Olsen RW, Cohen JB (2006) Identification of a GABAA receptor anesthetic binding site at subunit interfaces by photolabeling with an etomidate analog. J Neurosci 26(45):11599–11605

Lochner M, Lummis SC (2010) Agonists and antagonists bind to an AA interface in the heteromeric 5-HT3AB receptor. Biophys J 98(8):1494–1502

Lovinger DM (1991) Ethanol potentiation of 5-HT3 receptor-mediated ion current in NCB-20 neuroblastoma cells. Neurosci Lett 122(1):57–60

Lummis SC, Beene DL, Lee LW, Lester HA, Broadhurst RW, Dougherty DA (2005) Cis–trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. Nature 438(7065):248–252

Lummis SC, Thompson AJ, Bencherif M, Lester HA (2011) Varenicline is a potent agonist of the human 5-hydroxytryptamine3 receptor. J Pharmacol Exp Ther 339(1):125–131

Lummis SC (2012) 5-HT3 receptors. J Biol Chem 287(48):40239–40245

Lummis SC, Thompson AJ (2013) Agonists and antagonists induce different palonosetron dissociation rates in 5-HT3A and 5-HT3AB receptors. Neuropharmacology 73:241–246

Maricq AV, Peterson AS, Brake AJ, Myers RM, Julius D (1991) Primary structure and functional expression of the 5HT3 receptor, a serotonin-gated ion channel. Science 254(5030):432–437

McKinnon NK, Reeves DC, Akabas MH (2011) 5-HT3 receptor ion size selectivity is a property of the transmembrane channel, not the cytoplasmic vestibule portals. J Gen Physiol 138(4):453–466

McKinnon NK, Bali M, Akabas MH (2012) Length and amino acid sequence of peptides substituted for the 5-HT3A receptor M3M4 loop may affect channel expression and desensitization. PloS one 7(4)

Miller PS, Aricescu AR (2014) Crystal structure of a human GABA A receptor. Nature 512(7514):270–275

Miyake A, Mochizuki S, Takemoto Y, Akuzawa S (1995) Molecular cloning of human 5-hydroxytryptamine3 receptor: heterogeneity in distribution and function among species. Mol Pharmacol 48(3):407–416

Moore KA, Oh EJ, Weinreich D (2002) 5-HT3 receptors mediate inflammation-induced unmasking of functional tachykinin responses in vitro. J Appl Physiol 92(6):2529–2534

Morales-Perez CL, Noviello CM, Hibbs RE (2016) X-ray structure of the human α4β2 nicotinic receptor. Nature 538(7625):411

Mosesso R, Dougherty DA, Lummis SC (2019) Proline Residues in the transmembrane/extracellular domain interface loops have different behaviors in 5-HT3 and nACh receptors. ACS Chem Neurosci 10(7):3327–3333

Mott DD, Erreger K, Banke TG, Traynelis SF (2001) Open probability of homomeric murine 5-HT3A serotonin receptors depends on subunit occupancy. J Physiol 535(2):427–443

Mukhtasimova N, Sine SM (2013) Nicotinic receptor transduction zone: invariant arginine couples to multiple electron-rich residues. Biophys J 104(2):355–367

Mussa BM, Sartor DM, Verberne AJ (2008) Activation of cholecystokinin (CCK1) and serotonin (5-HT3) receptors increases the discharge of pancreatic vagal afferents. Eur J Pharmacol 601(1–3):198–206

Mylecharane E (1995) Ventral tegmental area 5-HT receptors: mesolimbic dopamine release and behavioural studies. Behav Brain Res 73(1–2):1–5

Niesler B, Frank B, Kapeller J, Rappold GA (2003) Cloning, physical mapping and expression analysis of the human 5-HT3 serotonin receptor-like genes HTR3C, HTR3D and HTR3E. Gene 310:101–111

Niesler B, Walstab J, Combrink S, Moeller D, Kapeller J, Rietdorf J, Bönisch H, Göthert M, Rappold G, Bruess M (2007) Characterization of the novel human serotonin receptor subunits 5-HT3C, 5-HT3D, and 5-HT3E. Mol Pharmacol 72(1):8–17

Noam Y, Wadman WJ, Van Hooft JA (2008) On the voltage-dependent Ca2+ block of serotonin 5-HT3 receptors: a critical role of intracellular phosphates. J Physiol 586(15):3629–3638

Nyce HL, Stober ST, Abrams CF, White MM (2010) Mapping spatial relationships between residues in the ligand-binding domain of the 5-HT3 receptor using a molecular ruler. Biophys J 98(9):1847–1855

Oostland M, Buijink MR, van Hooft JA (2013) Serotonergic control of Purkinje cell maturation and climbing fibre elimination by 5-HT3 receptors in the juvenile mouse cerebellum. J Physiol 591(7):1793–1807

Panicker S, Cruz H, Arrabit C, Slesinger PA (2002) Evidence for a centrally located gate in the pore of a serotonin-gated ion channel. J Neurosci 22(5):1629–1639

Parker RM, Barnes JM, Ge J, Barber PC, Barnes NM (1996) Autoradiographic distribution of [3H]-(S)-zacopride-labelled 5-HT3 receptors in human brain. J Neurol Sci 144(1–2):119–127

Paulsen IM, Martin IL, Dunn SM (2009) Isomerization of the proline in the M2–M3 linker is not required for activation of the human 5-HT3A receptor. J Neurochem 110(3):870–878

Peters JA, Hales TG, Jeremy JL (2005) Molecular determinants of single-channel conductance and ion selectivity in the cys-loop family: insights from the 5-HT3 receptor. Trends Pharmacol Sci 26(11):587–594

Phulera S, Zhu H, Yu J, Claxton DP, Yoder N, Yoshioka C, Gouaux E (2018) Cryo-EM structure of the benzodiazepine-sensitive $\alpha 1\beta 1\gamma 2S$ tri-heteromeric GABAA receptor in complex with GABA. Elife 7:e39383

Pless SA, Lynch JW (2009) Magnitude of a conformational change in the glycine receptor β1-β2 loop is correlated with agonist efficacy. J Biol Chem 284(40):27370–27376

Polovinkin L, Hassaine G, Perot J, Neumann E, Jensen AA, Lefebvre SN, Corringer P-J, Neyton J, Chipot C, Dehez F (2018) Conformational transitions of the serotonin 5-HT 3 receptor. Nature 563(7730):275–279

Powley TL, Spaulding RA, Haglof SA (2011) Vagal afferent innervation of the proximal gastrointestinal tract mucosa: chemoreceptor and mechanoreceptor architecture. J Comp Neurol 519(4):644–660

Price KL, Lummis SC (2004) The role of tyrosine residues in the extracellular domain of the 5-hydroxytryptamine3 receptor. J Biol Chem 279(22):23294–23301

Price KL, Millen KS, Lummis SC (2007) Transducing agonist binding to channel gating involves different interactions in 5-HT3 and GABAC receptors. J Biol Chem 282(35):25623–25630

Price KL, Bower KS, Thompson AJ, Lester HA, Dougherty DA, Lummis SC (2008) A hydrogen bond in loop A is critical for the binding and function of the 5-HT3 receptor. Biochemistry 47(24):6370–6377

Puig MV, Santana N, Celada P, Mengod G, Artigas F (2004) In vivo excitation of GABA interneurons in the medial prefrontal cortex through 5-HT3 receptors. Cereb Cortex 14(12):1365–1375

Purohit P, Auerbach A (2013) Loop C and the mechanism of acetylcholine receptor–channel gating. J Gen Physiol 141(4):467–478

Rácz I, Bilkei-Gorzo A, Markert A, Stamer F, Göthert M, Zimmer A (2008) Anandamide effects on 5-HT3 receptors in vivo. Eur J Pharmacol 596(1–3):98–101

Rajkumar R, Mahesh R (2010) The auspicious role of the 5-HT3 receptor in depression: a probable neuronal target? J Psychopharmacol 24(4):455–469

Rammes G, Eisensamer B, Ferrari U, Shapa M, Gimpl G, Gilling K, Parsons C, Riering K, Hapfelmeier G, Bondy B (2004) Antipsychotic drugs antagonize human serotonin type 3 receptor currents in a noncompetitive manner. Mol Psychiatry 9(9):846–858

Raybould HE, Glatzle J, Robin C, Meyer JH, Phan T, Wong H, Sternini C (2003) Expression of 5-HT3 receptors by extrinsic duodenal afferents contribute to intestinal inhibition of gastric emptying. Am J Physiol-Gastrointest Liver Physiol 284(3):G367–G372

Rayes D, De Rosa MJ, Sine SM, Bouzat C (2009) Number and locations of agonist binding sites required to activate homomeric Cys-loop receptors. J Neurosci 29(18):6022–6032

Reeves DC, Goren EN, Akabas MH, Lummis SC (2001) Structural and electrostatic properties of the 5-HT3Receptor pore revealed by substituted cysteine accessibility mutagenesis. J Biol Chem 276(45):42035–42042

Reeves DC, Sayed MF, Chau P-L, Price KL, Lummis SC (2003) Prediction of 5-HT3 receptor agonist-binding residues using homology modeling. Biophys J 84(4):2338–2344

Reeves DC, Jansen M, Bali M, Lemster T, Akabas MH (2005) A role for the β1-β2 loop in the gating of 5-HT3 receptors. J Neurosci 25(41):9358–9366

Revah F, Bertrand D, Galzi J-L, Devillers-Thiery A, Mulle C, Hussy N, Bertrand S, Ballivet M, Changeux J-P (1991) Mutations in the channel domain alter desensitization of a neuronal nicotinic receptor. Nature 353(6347):846–849

Roberts A, Grafton G, Powell AD, Brock K, Chen C, Xie D, Huang J, Liu S, Cooper AJ, Brady CA (2020) CSTI-300; a novel 5-HT3 receptor partial agonist with potential to treat patients with irritable bowel syndrome or carcinoid syndrome. J Pharmacol Exp Ther

Roerig B, Nelson DA, Katz LC (1997) Fast synaptic signaling by nicotinic acetylcholine and serotonin 5-HT3 receptors in developing visual cortex. J Neurosci 17(21):8353–8362

Rojas C, Slusher BS (2012) Pharmacological mechanisms of 5-HT3 and tachykinin NK1 receptor antagonism to prevent chemotherapy-induced nausea and vomiting. Eur J Pharmacol 684(1–3):1–7

Rondé P, Nichols RA (1998) High calcium permeability of serotonin 5-HT3 receptors on presynaptic nerve terminals from rat striatum. J Neurochem 70(3):1094–1103

Rosenberg M, Pie B, Cooper E (1997) Developing neonatal rat sympathetic and sensory neurons differ in their regulation of 5-HT3 receptor expression. J Neurosci 17(17):6629–6638

Rüsch D, Musset B, Wulf H, Schuster A, Raines DE (2007) Subunit-dependent modulation of the 5-hydroxytryptamine type 3 receptor open-close equilibrium by n-alcohols. J Pharmacol Exp Ther 321(3):1069–1074

Sakmann B, Patlak J, Neher E (1980) Single acetylcholine-activated channels show burst-kinetics in presence of desensitizing concentrations of agonist. Nature 286(5768):71–73

Schofield PR, Darlison MG, Fujita N, Burt DR, Stephenson FA, Rodriguez H, Rhee LM, Ramachandran J, Reale V, Glencorse TA (1987) Sequence and functional expression of the GABAA receptor shows a ligand-gated receptor super-family. Nature 328(6127):221

Sevoz C, Callera J, Machado B, Hamon M, Laguzzi R (1997) Role of serotonin3 receptors in the nucleus tractus solitarii on the carotid chemoreflex. Am J Physiol-Heart Circulatory Physiol 272(3):H1250–H1259

Solt K, Stevens RJ, Davies PA, Raines DE (2005) General anesthetic-induced channel gating enhancement of 5-hydroxytryptamine type 3 receptors depends on receptor subunit composition. J Pharmacol Exp Ther 315(2):771–776

Solt K, Ruesch D, Forman SA, Davies PA, Raines DE (2007) Differential effects of serotonin and dopamine on human 5-HT3A receptor kinetics: interpretation within an allosteric kinetic model. J Neurosci 27(48):13151–13160

Spier AD, Lummis SC (2000) The role of tryptophan residues in the 5-Hydroxytryptamine3 receptor ligand binding domain. J Biol Chem 275(8):5620–5625

Steward LJ, Boess FG, Steele JA, Liu D, Wong N, Martin IL (2000) Importance of phenylalanine 107 in agonist recognition by the 5-hydroxytryptamine3A receptor. Mol Pharmacol 57(6):1249–1255

Sugita S, Shen K-Z, North R (1992) 5-hydroxytryptamine is a fast excitatory transmitter at 5-HT3 receptors in rat amygdala. Neuron 8(1):199–203

Sullivan NL, Sullivan NL, Thompson AJ, Price KL, Lummis SC (2006) Defining the roles of Asn-128, Glu-129 and Phe-130 in loop A of the 5-HT3 receptor. Mol Membr Biol 23(5):442–451

Terejko K, Kaczor PT, Michałowski MA, Dąbrowska A, Mozrzymas JW (2020) The C loop at the orthosteric binding site is critically involved in GABAA receptor gating. Neuropharmacology 166:107903

Thompson AJ, Lummis SC (2003) A single ring of charged amino acids at one end of the pore can control ion selectivity in the 5-HT3 receptor. Br J Pharmacol 140(2):359–365

Thompson AJ, Padgett CL, Lummis SC (2006) Mutagenesis and molecular modeling reveal the importance of the 5-HT3 receptor F-loop. J Biol Chem 281(24):16576–16582

Thompson AJ, Sullivan NL, Lummis SC (2006) Characterization of 5-HT 3 receptor mutations identified in schizophrenic patients. J Mol Neurosci 30(3):273–281

Thompson AJ, Lochner M, Lummis S (2008) Loop B is a major structural component of the 5-HT3 receptor. Biophys J 95(12):5728–5736

Thompson A, Lummis S (2008) Antimalarial drugs inhibit human 5-HT3 and GABAA but not GABAC receptors. Br J Pharmacol 153(8):1686–1696

Thompson AJ, Duke RK, Lummis SC (2011) Binding sites for bilobalide, diltiazem, ginkgolide, and picrotoxinin at the 5-HT3 receptor. Mol Pharmacol 80(1):183–190

Thompson A, Price K, Lummis S (2011) Cysteine modification reveals which subunits form the ligand binding site in human heteromeric 5-HT3AB receptors. J Physiol 589(17):4243–4257

Thompson AJ (2013) Recent developments in 5-HT3 receptor pharmacology. Trends Pharmacol Sci 34(2):100–109

Tzvetkov MV, Meineke C, Oetjen E, Hirsch-Ernst K, Brockmöller J (2007) Tissue-specific alternative promoters of the serotonin receptor gene HTR3B in human brain and intestine. Gene 386(1–2):52–62

Unwin N (2005) Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. J Mol Biol 346(4):967–989

Venkataraman P, Venkatachalan SP, Joshi PR, Muthalagi M, Schulte MK (2002) Identification of critical residues in loop E in the 5-HT 3AS R binding site. BMC Biochem 3(1):15

Walsh RM, Roh S-H, Gharpure A, Morales-Perez CL, Teng J, Hibbs RE (2018) Structural principles of distinct assemblies of the human α4β2 nicotinic receptor. Nature 557(7704):261–265

Walstab J, Wohlfarth C, Hovius R, Schmitteckert S, Röth R, Lasitschka F, Wink M, Bönisch H, Niesler B (2014) Natural compounds boldine and menthol are antagonists of human 5-HT 3 receptors: implications for treating gastrointestinal disorders. Neurogastroenterol Motil 26(6):810–820

Wan S, Browning KN (2008) Glucose increases synaptic transmission from vagal afferent central nerve terminals *via* modulation of 5-HT3 receptors. Am J Physiol-Gastrointest Liver Physiol 295(5):G1050–G1057

Wang H-L, Cheng X, Taylor P, McCammon JA, Sine SM (2008) Control of cation permeation through the nicotinic receptor channel. PLoS Comput Biol 4(2)

Wong E, Clark R, Leung E, Loury D, Bonhaus D, Jakeman L, Parnes H, Whiting R, Eglen R (1995) The interaction of RS 25259–197, a potent and selective antagonist, with 5-HT3 receptors, in vitro. Br J Pharmacol 114(4):851–859

Wozniak KM, Pert A, Linnoila M (1990) Antagonism of 5-HT3 receptors attenuates the effects of ethanol on extracellular dopamine. Eur J Pharmacol 187(2):287–289

Wu Q, Clark MS, Palmiter RD (2012) Deciphering a neuronal circuit that mediates appetite. Nature 483(7391):594–597

Xiong W, Hosoi M, Koo B-N, Zhang L (2008) Anandamide inhibition of 5-HT3A receptors varies with receptor density and desensitization. Mol Pharmacol 73(2):314–322

Xiong W, Cheng K, Cui T, Godlewskirole of loop F residues in determining differential G, Rice KC, Xu Y, Zhang L (2011) Cannabinoid potentiation of glycine receptors contributes to cannabis-induced analgesia. Nat Chem Biol 7(5):296

Xiu X, Hanek AP, Wang J, Lester HA, Dougherty DA (2005) A unified view of the role of electrostatic interactions in modulating the gating of Cys loop receptors. J Biol Chem 280(50):41655–41666

Yakel J, Lagrutta A, Adelman J, North R (1993) Single amino acid substitution affects desensitization of the 5-hydroxytryptamine type 3 receptor expressed in Xenopus oocytes. Proc Natl Acad Sci 90(11):5030–5033

Yan D, Schulte MK, Bloom KE, White MM (1999) Structural features of the ligand-binding domain of the serotonin 5HT3 receptor. J Biol Chem 274(9):5537–5541

Yang J (1990) Ion permeation through 5-hydroxytryptamine-gated channels in neuroblastoma N18 cells. J Gen Physiol 96(6):1177–1198

Yang J, Mathie A, Hille B (1992) 5-HT3 receptor channels in dissociated rat superior cervical ganglion neurons. J Physiol 448(1):237–256

Yang K-H, Galadari S, Isaev D, Petroianu G, Shippenberg TS, Oz M (2010) The nonpsychoactive cannabinoid cannabidiol inhibits 5-hydroxytryptamine3A receptor-mediated currents in Xenopus laevis oocytes. J Pharmacol Exp Ther 333(2):547–554

Yuan S, Filipek S, Vogel H (2016) A gating mechanism of the serotonin 5-HT3 receptor. Structure 24(5):816–825

Zarkadas E, Zhang H, Cai W, Effantin G, Perot J, Neyton J, Chipot C, Schoehn G, Dehez F, Nury H (2020) The binding of palonosetron and other antiemetic drugs to the serotonin 5-HT3 receptor. BioRxiv

Zhang R, Wen X, Militante J, Hester B, Rhubottom HE, Sun H, Leidenheimer NJ, Yan D, White MM, Machu TK (2007) The role of loop F residues in determining differential d-tubocurarine potencies in mouse and human 5-hydroxytryptamine 3A receptors. Biochemistry 46(5):1194–1204

Zhang QJ, Li LB, Niu XL, Liu J, Gui ZH, Feng JJ, Ali U, Hui YP, Wu ZH (2011) The pyramidal neurons in the medial prefrontal cortex show decreased response to 5-hydroxytryptamine-3 receptor stimulation in a rodent model of Parkinson's disease. Brain Res 1384:69–79

Zheng Y, Yu T, Tang Y, Xiong W, Shen X, Jiang L, Lin L (2017) Efficacy and safety of 5-hydroxytryptamine 3 receptor antagonists in irritable bowel syndrome: a systematic review and meta-analysis of randomized controlled trials. PLoS One 12(3)

Zhu S, Noviello CM, Teng J, Walsh RM, Kim JJ, Hibbs RE (2018) Structure of a human synaptic GABA A receptor. Nature 559(7712):67–72

Ziemba PM, Schreiner BS, Flegel C, Herbrechter R, Stark TD, Hofmann T, Hatt H, Werner M, Gisselmann G (2015) Activation and modulation of recombinantly expressed serotonin receptor type 3A by terpenes and pungent substances. Biochem Biophys Res Commun 467(4):1090–1096

Chapter 12
The SF3b Complex is an Integral Component of the Spliceosome and Targeted by Natural Product-Based Inhibitors

Nicholas A. Larsen

Abstract In this chapter, the essential role of the SF3b multi-protein complex will be discussed in the context of the overall spliceosome. SF3b is critical during spliceosome assembly for recognition of the branch point (BP) adenosine and, by de facto, selection of the 3′ splice site. This complex is highly dynamic, undergoing significant conformational changes upon loading of the branch duplex RNA and in its relative positioning during spliceosomal remodeling from the A, pre-B, B, B^{act} and B* complexes. Ultimately, during the spliceosome activation phase, SF3b must be displaced to unmask the branch point adenosine for the first splicing reaction to occur. In certain cancers, such as the hematological malignancies CML, CLL and MDS, the SF3b subunit SF3B1 is frequently mutated. Recent studies suggest these mutations lead to inappropriate branch point selection and mis-splicing events that appear to be drivers of disease. Finally, the SF3b complex is the target for at least three different classes of natural product-based inhibitors. These inhibitors bind in the BP adenosine-binding pocket and demonstrate a pre-mRNA competitive mechanism of action resulting in either intron retention or exon skipping. These compounds are extremely useful as chemical probes to isolate and characterize early stages of spliceosome assembly. They are also being explored preclinically and clinically as possible agents for hematological cancers.

Keywords RNA splicing · SF3B1 · PHF5a · Pladienolide · SF3b · U2 snRNP

Abbreviations

BP	Branch point
BPRS	Branch point recognition sequence
BS	Branch sequence
BSL	Branch stem loop

N. A. Larsen (✉)
H3 Biomedicine, Inc., Cambridge, MA 02139, USA
e-mail: nicholas_larsen@h3biomedicine.com

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_12

CMML	Chronic myelomonocytic leukemia
CLL	Chronic lymphocytic leukemia
EM	Electron microscopy
ILS	Intron lariat spliceosome
ISL	Internal stem loop
MDS	Myelodysplastic syndromes
mRNA	Messenger RNA
NMD	Nonsense-mediated decay
NMR	Nuclear magnetic resonance
NTC	Nineteen complex
NTR	Nineteen related
Plad	Pladienolide
PY	Polypyrimidine
RNA	Ribonucleic acid
RRM	RNA recognition motif
SF	Splicing factor
snRNP	Small nuclear ribonucleoproteins
snRNA	Small nuclear RNAs
SS	Splice site
ULM	U2AF ligand motifs
UHM	U2AF homology motif
U2AF	U2 auxiliary factor
WT	Wild type

Introduction

The spliceosome is responsible for the removal of introns and ligation of exons to form mature mRNA. Introns vary dramatically in both sequence and length, which poses a fascinating molecular recognition problem. Namely, how can the spliceosome catalyze a reaction involving a vast diversity of pre-mRNA substrates? To answer this, over the past 20+ years, X-ray crystallography and NMR techniques have provided atomic-resolution snapshots for many of the individual components. At the same time, myriads of elegant biochemical and yeast genetic studies have provided incredible insight into their function. Now, thanks to the resolution revolution in cryo-EM, the global context has been elucidated for these previous discoveries, enabling a truly integrated understanding of spliceosome structure and function. Still, there is much more to learn. The cryo-EM structures have revealed numerous protein–protein and protein-RNA interactions previously unknown or unappreciated. Functional relationships can be deduced, but additional biochemistry, in the years to come, is necessary to further understand this trove of structural data.

At a high level, we have learned that the spliceosome, like the ribosome, is a complex ribozyme. Numerous small nuclear ribonucleoproteins (snRNPs) assemble

in an external framework around the catalytic small nuclear RNAs (snRNAs) to choreograph splicing. Several excellent overviews have been written on this topic (Kastner et al. 2019; Wan et al. 2019b; Wilkinson et al. 2019). Here, we provide a brief synopsis, which will serve as a framework for better understanding the role of the SF3b subcomplex. Then, we will review the role of the SF3b complex in terms of its structure and function, its role in disease, and finally as a target for drug discovery.

Splicing Overview

The spliceosome catalyzes two transesterification reactions, the branching reaction and the exon-ligation reaction (Fig. 12.1a). In the course of these two reactions, the spliceosome undergoes four broadly defined stages including (1) assembly, (2) activation, (3) catalysis, and (4) disassembly (Fig. 12.1b). In the first assembly stage, the spliceosome must recognize elements within the intron. Each intron is flanked by the 5′ and 3′ splice sites (5′ SS and 3′ SS). The 5′ SS contains the conserved GU sequence, and the 3′ SS contains a conserved AG. Upstream of the 3′ SS may be found a tract of 10–15 polypyrimidines (PYs) followed by the branch sequence (BS), which surrounds the branch point adenosine (Fig. 12.1a). The branch point adenosine initiates nucleophilic attack at the 5′ SS during the first branching reaction and therefore is a central player in splicing catalysis (Fig. 12.1a).

During the initial assembly phase of the spliceosome, these conserved sequences in the intron are marked by different proteins (Fig. 12.1b). The 5′ SS is marked by the U1 snRNP via base pairing from the 5′ end of the U1 snRNA (Pomeranz et al. 2009; Kondo et al. 2015; Li et al. 2019). The U1 snRNP is composed of the seven Sm proteins that form a ring. The 3′ end of the U1 snRNA threads through this ring and then forms a terminal stem loop, seemingly holding the protein complex in place like a washer on a bolt. (The Sm proteins serve a similar structural role on the other snRNA's U2, U4, and U5 with a related LSm complex encircling the 3′ end of U6.) The U1 snRNP also contains three additional proteins (U1-70, U1A, and U1C). U1-70 and U1A contain RRM domains that bind to stem loops in the U1 snRNA and help strengthen interactions with the 5′ SS, while U1C contains a zinc finger motif that stabilizes the duplex between the 5′ SS and U1 snRNA (Kondo et al. 2015; Li et al. 2019).

Meanwhile, on the 3′ side of the intron, the branch point adenosine and polypyrimidine (PY) tract are first recognized by splicing factor 1 (SF1) and U2 auxiliary factors U2AF2/1, respectively (Fig. 12.1b). The NMR structure of SF1 bound to the branch point sequence beautifully illustrates how it selectively binds adenosine through pseudo base-pairing interactions to the protein backbone, which precludes binding of other nucleotides (Liu et al. 2001). Additional cooperative interactions between SF1 and conserved bases in the BS further define selectivity for the BP adenosine. At the same time, the neighboring polypyrimidine (PY) tract and 3′ SS are recognized by U2AF2 and U2AF1, respectively (Wang et al. 2013; Agrawal et al. 2016). This

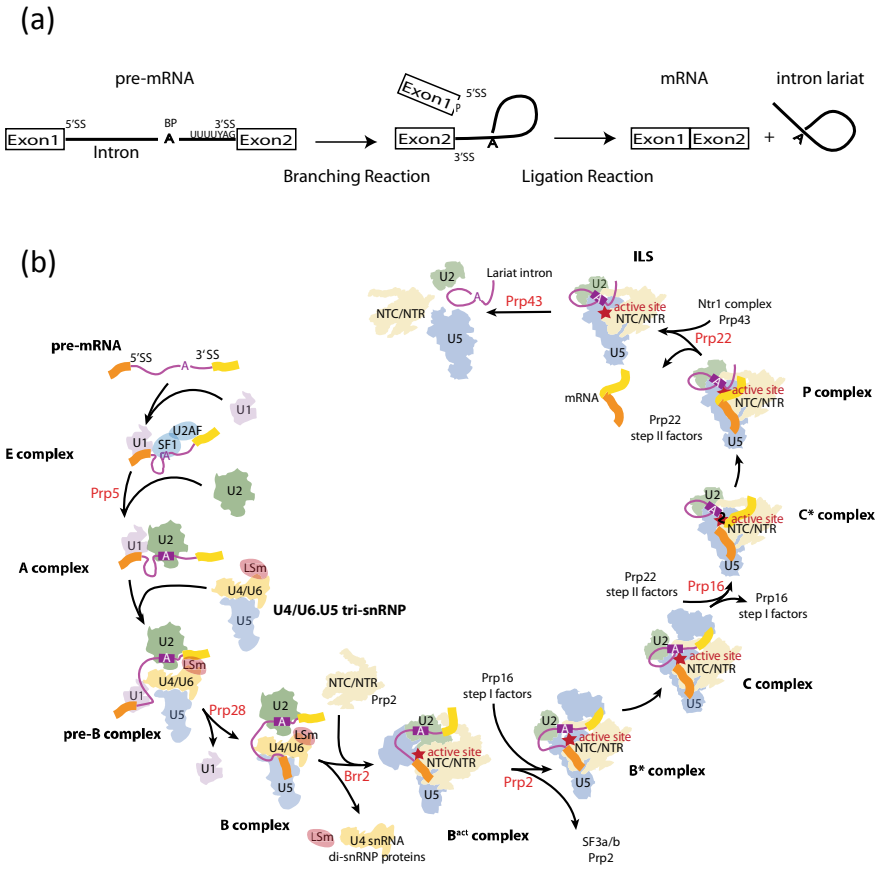


Fig. 12.1 Overview of splicing. **a** Schematic showing the basic layout of the pre-mRNA with the 5′ and 3′ splice sites (SS) and the branch point adenosine (BP). In the branching reaction, the BP attacks the 5′ SS liberating exon 1. In the ligation reaction, the 5′ SS attacks the 3′ SS to form the final mRNA and liberate the intron lariat sequence (ILS). **b** Overview of all the components and remodeling steps involved in spliceosome assembly, activation, catalysis, and disassembly. Note the intron has been rendered in magenta color scheme, which will be followed in subsequent figures. Adapted with kind permission from M.E. Wilkinson

first stage of spliceosome assembly of the U1snRNP at the 5′ SS and SF1 and U2AFs at the 3′ SS is referred to as the "E complex" (Fig. 12.1b).

In the next step of assembly, the first ATP-dependent step occurs in which the U2 snRNP replaces SF1 and U2AF2/1 at the 3′ SS. Incorporation of the U2 snRNP and departure of SF1 and U2AF is triggered by the DEAD-box helicase Prp5 (Xu et al. 2004). Prp5 facilitates formation of the branch point RNA duplex between the U2 snRNA and the pre-mRNA branch point sequence and loading of the U2 snRNP at the 3′ SS. This stage, with the U1 and U2 snRNPs bound at the 5′ and 3′ sites respectively, is called the "A complex" (Fig. 12.1b), and a cryo-EM structure from Saccharomyces

cerevisiae has been determined at 4.0 Å (Plaschka et al. 2018). Compositionally, as previously mentioned, the U2 snRNP also contains a SM ring complex near the 3′ end of the U2 snRNA (Fig. 12.2a, b). In addition, it has two bridging proteins, U2A′ and U2B″, and the SF3a and SF3b subcomplexes, each consisting of three and seven additional proteins respectively (Fig. 12.2a, b). The SF3b complex binds to the branch point duplex (also known as branch helix) with the unpaired branch point adenosine bulged out and nestled in a pocket formed between SF3b proteins SF3B1 and PHF5a (Fig. 12.2c). Unlike the structure of SF1, which shows specific interactions to the base that are critical for branch point selectivity, the SF3B1-PHF5A binding pocket does not form any H-bond interactions with the base (Fig. 12.2c). Instead, selectivity is defined by base-pairing interactions in the branch helix, while the SF3ba/b subcomplexes form a protein shell or chaperone-like structure around the branch helix and BP adenosine, forming electrostatic interactions primarily with the phosphate backbone (Fig. 12.2c). The SF3b subcomplex will be discussed in further detail in subsequent sections.

Once the "A complex" has formed, the next phase of spliceosomal assembly can occur. At this point, the preassembled tri-snRNP, consisting of U4, U5, and U6, is recruited to form the "pre-B complex" (Figs. 12.1b and 12.3a) (Bai et al. 2018; Charenton et al. 2019). Here, we see the first base pairing interactions formed between the 5′ end of the U2 and 3′ end of the U6 snRNAs. In this pre-B complex, the U1 snRNP retains the 5′ SS, and the U6 snRNA is masked by the U4 snRNA to maintain this pre-catalytic assembly state. Eventually, U6 will become the catalytic core of the spliceosome, interacting with two metal ions that catalyze the branching and exon-ligation reactions. In addition to the snRNA's, several key proteins are associated with the tri-snRNP and incorporated into the pre-B complex. These include helicases Prp28, Brr2, Prp8, and the GTPase Snu114. At this stage, Brr2 forms loose interactions with one of the seven proteins in the SF3b complex, SF3B3.

The transition from the pre-B to B complex requires two important transformations (Fig. 12.1b) (Bertram et al. 2017a; Plaschka et al. 2017). Firstly, the helicase Prp28 dislodges the U1 snRNP so that the 5′ exon base pairs with the first stem loop of U5 snRNA and the 5′ SS base pairs with the U6 ACAGAGA box loop. These are key transformational steps in transferring the 5′ SS to what will become the catalytic center (Newman and Norman 1992; Sontheimer and Steitz 1993). Subsequently, additional remodeling occurs in which the helicase Prp8 RNaseH domain pivots and Brr2 loads onto the U4 snRNA. This loading step poises Brr2 to unwind U4 from U6, which is critical for activation.

The first step in activation occurs in the formation of the B^{act} complex (Fig. 12.1b) (Rauhut et al. 2016; Yan et al. 2016; Haselbach et al. 2018; Zhang et al. 2018). Here, Brr2 unwinds the U4/U6 snRNA helix, facilitating the formation of more extensive pairing between U2 and U6 as well as the formation of the U6 internal stem loop (ISL) in the catalytic core of the spliceosome. The ISL now binds the two Mg ions that are important for stabilizing the leaving groups during the branching and exon-ligation reactions. In this dramatic rearrangement, the U4 snRNA departs from the spliceosome and two new complexes are recruited, the so-called nineteen complex (NTC) and nineteen related (NTR) complexes (Plaschka et al. 2017) (Fig. 12.1b).

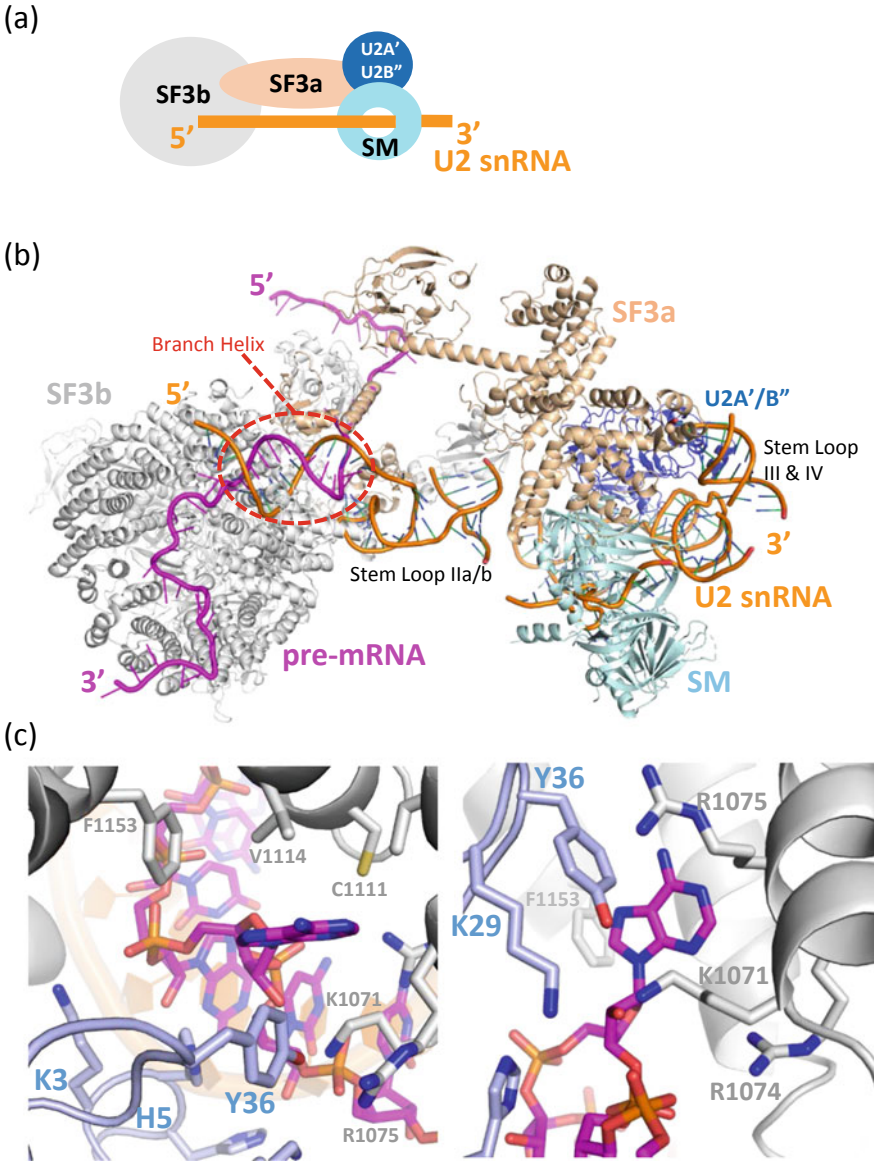


Fig. 12.2 Overview of the U2 snRNP bound to pre-mRNA. **a** Two dimensional schematic of the 17S U2 snRNP and **b** cryo-EM structure showing the location of SF3b, SF3a, U2A′ & U2B″, and the SM ring relative to each other and the pre-mRNA and U2 snRNA. The branch helix is indicated by the red dashed ellipse. The coloring scheme is the same for (**a**) and (**b**). **c** Top and side views of the branch point adenosine sandwiched in the interface between SF3B1 (gray) and PHF5a (blue). Interactions with adenine are hydrophobic while numerous electrostatic and H-bond interactions are formed with the RNA phosphate backbone

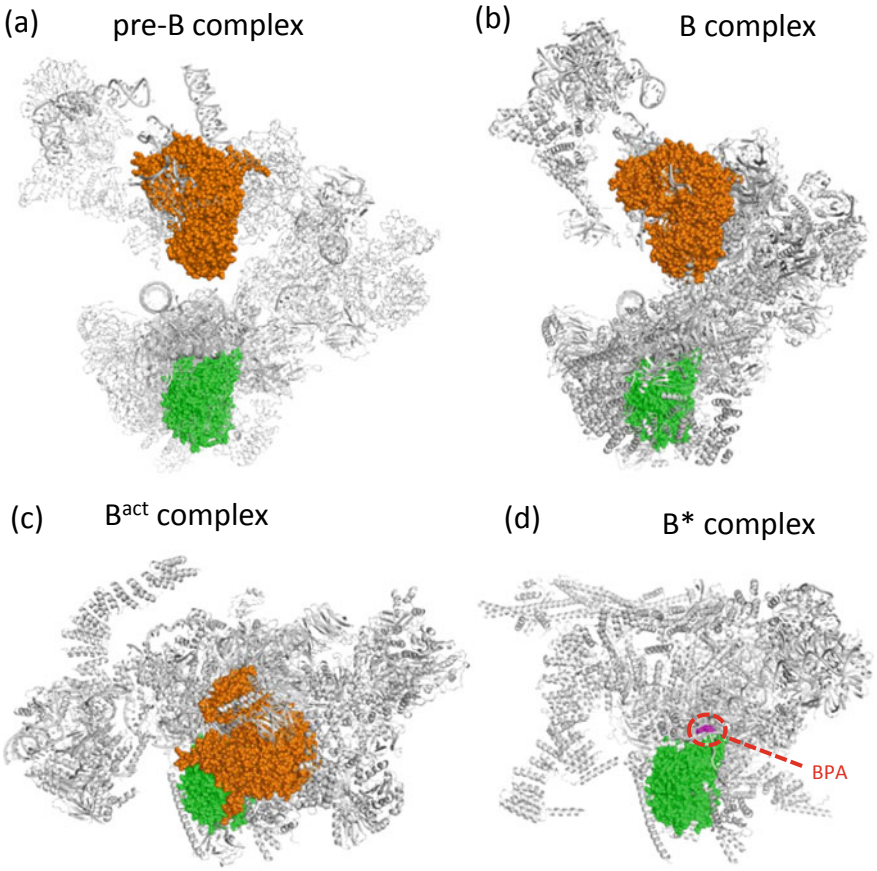


Fig. 12.3 Relative position of SF3B1 and PHF5a (both shown in orange) in the spliceosome. Each image is generated from the available cryo-EM structure from yeast. All views are from the same orientation using the reverse transcriptase homology domain of PRP8 (green) as a reference frame. The most dramatic remodeling occurs from B to B^{act}. Subsequently, the entire SF3b complex is removed from the spliceosome in the remodeling step from B^{act} to B*. Removal of SF3b exposes the branch point adenosine (BPA) for the first branching reaction

The removal of U4 also allows the entire U2 snRNP to swing down, bringing the branch point adenosine into closer proximity to the catalytic center (Fig. 12.3a–c). Throughout these first few steps, however, the branch point adenosine is sequestered away from the active site by the SF3b sub-complex (Fig. 12.1a) (Yan et al. 2016).

Up to this point, the branch point adenosine has been selectively stabilized by the branch duplex, which, in turn, is sequestered by the SF3b complex. Specifically, the branch point adenosine is held between SF3B1 and PHF5a and therefore is sterically occluded some 50 Å removed from the 5′ SS and the catalytic center (Fig. 12.3c, d). In order for the branching reaction to occur, the SF3a/b subcomplexes must depart,

which is facilitated by the helicase Prp2 (Lardelli et al. 2010) (Fig. 12.1b). Prp2 binds downstream of the branch point and disrupts the interactions between SF3a/b and the branch helix. This activation stage, defined by the departure of SF3b sub-complex and translocation of the branch point into proximity of the catalytic center, is referred to as the B* complex (Fig. 12.1b) (Wan et al. 2019a).

In the B* complex, the branch point adenosine is still some 4 Å away from the reactive center. In the first stage of catalysis, additional splicing factor proteins Yju2 and Cwc25 help stabilize the branch point for the first nucleophilic attack of the branching reaction resulting in the intron lariat (Fig. 12.1a, b). This structure is referred to as the "C complex" (Galej et al. 2016; Wan et al. 2016; Zhan et al. 2018) (Fig. 12.1b). At this point, the branch helix and branching factors Yju2 and Cwc25 sterically occlude the active site and must be moved aside to make room for the 3′ SS. This remodeling is accomplished by the helicase Prp16 (Tseng et al. 2011). Prp16 displaces the branching factors Yju2 and Cwc25 and pivots the branch helix to make room for the 3′ SS. Now, with the assistance of exon-ligation factors Prp18 and Slu7 (Ohrt et al. 2013), the 3′ SS can stack on the 5′ SS-U6 snRNA ACAGAGA box loop pair and be poised for the final exon-ligation reaction (Shi 2017). Nearly all 3′ SS's contain the sequence YAG where Y is a pyrimidine base. Since this motif is commonly found, the 3′ SS is generally, by default, the first such YAG motif occurring within 10 nucleotides of the BP adenosine (Horowitz 2012). This means that the fidelity of exon ligation depends largely on selection of the BP adenosine, which becomes the de facto determinant for the 3′ SS. This spliceosome conformation with the 3′ SS poised in the catalytic center prior to ligation is referred to as the "C* complex" (Bertram et al. 2017b; Fica et al. 2017; Yan et al. 2017; Zhang et al. 2017) and after exon ligation as the post-catalytic "P complex" (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017; Fica et al. 2019) (Fig. 12.1b).

During the last disassembly phase of splicing, additional transformations occur (Fig. 12.1b). In the first stage of disassembly, the mRNA product is released via the action of Prp22, resulting in the intron lariat spliceosome or "ILS complex" (Hang et al. 2015; Yan et al. 2015; Wan et al. 2017). Prp22 interacts with RNA upstream of the 3′ SS prior to ligation and downstream of the 3′ SS after ligation. It is thought that translocation of Prp22 along the RNA leads to release of mRNA product. Prp22 also facilitates departure of exon-ligation factors Prp18 and Slu7 (Chen et al. 2013). Because the intron lariat is still buried in the heart of the spliceosome, sandwiched between the NTR and NTC complexes, additional disassembly must occur to allow for recycling of the splicing factors and eventual nonsense-mediated decay of the intron lariat. Therefore, the splicing factor Prp43 is recruited, and this helicase completes the disassembly through release of the intron lariat, U2 snRNP core, U5 snRNP, U6 snRNA, and NTC/R proteins (Toroney et al. 2019) (Fig. 12.1b).

U2 snRNP and the SF3b Subcomplex

The U2 snRNP consists of several protein sub-complexes that surround and chaperone the U2 snRNA. In humans, this snRNA contains four stem loops (Fig. 12.4a) (van der Feltz and Hoskins 2019). As previously discussed, the first proximal 5′ stem loop will eventually unwind during the transition from the B to B^{act} complexes and pair with U6 snRNA in the catalytic center of the spliceosome. The SM ring and the U2A′ and U2B″ proteins are located towards the more distal 3′ periphery between stem loops II and III (Fig. 12.2a). This minimal U2 complex with the U2 snRNA, U2A′, U2B″, and the SM ring is referred to as the 12S core particle and remains together throughout splicing. Between stem loop I and II is the critical

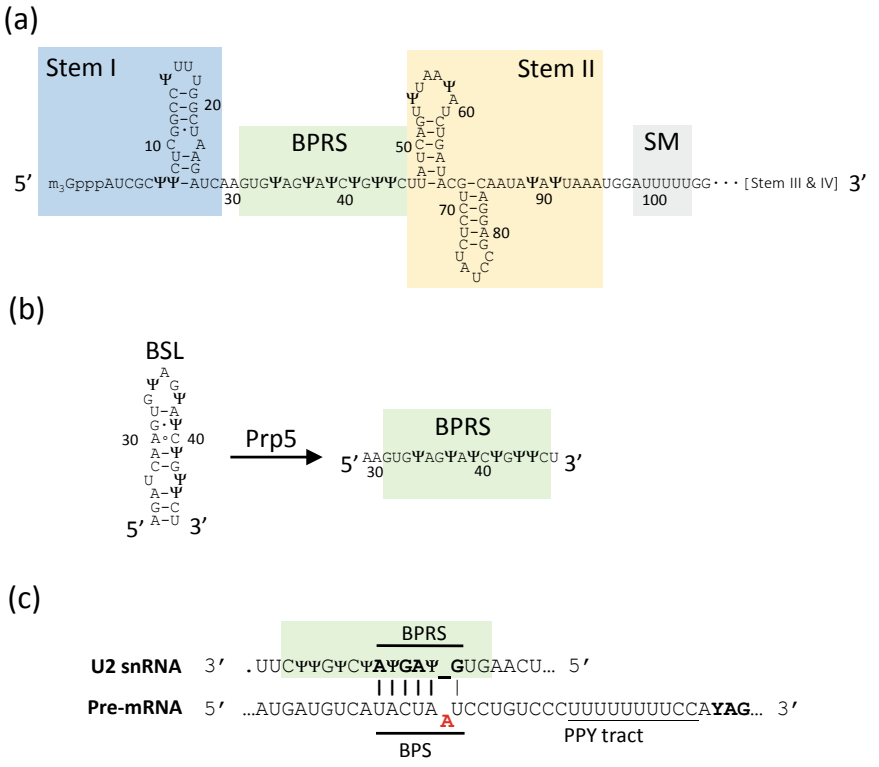


Fig. 12.4 Schematic of U2 snRNA. **a** Human U2 snRNA. The stem loops, branch point recognition sequence (BPRS), and SM binding sites are indicated. Ψ denotes a pseudouridine base. **b** In yeast, the BPRS is masked by an internal branch stem loop (BSL) before annealing to the pre-mRNA. Given that the corresponding human sequence in this region is identical, the BSL likely occurs in humans as well. Prp5 is believed to unwind the BSL to reveal the BPRS. **c** Typical interactions drawn between the BPRS (now drawn 3′–5′) and a prototypic pre-mRNA substrate (drawn 5′–3′) that make up the branch helix. The unpaired branch point adenosine is highlighted in red font. Prp5 facilitates the unwinding of the BSL and formation of the branch helix

branch point recognition sequence (BPRS) that is critical for pairing with the intron branch sequence (BS), forming the branch helix, and in this process, singling out the branch point adenosine nucleophile (Fig. 12.4a–c). Prior to branch helix formation, it has been shown, in yeast, that the U2 snRNA BPRS region forms an internal branch stem loop (BSL) that masks this region prior to association with the pre-mRNA (Fig. 12.4b) (Perriman and Ares 2010). The helicase Prp5 is necessary to unwind this branch stem loop to facilitate pairing with the intron branch site and formation of the branch helix (Perriman and Ares 2010).

In addition to the stable core subunits, the U2 snRNP contains the dynamically associated sub-complexes SF3a and SF3b (Fig. 12.2a–b) (Brosi et al. 1993). SF3a consists of 3 proteins (SF3a60, SF3a66, and SF3a120, also known as Prp9, Prp11 and Prp21, respectively, in yeast) (Brosi et al. 1993), and the SF3b complex contains 7 proteins (SF3B1, SF3B2, SF3B3, SF3B4, SF3B5, P14, and PHF5a) (Golas et al. 2003). SF3b associates with the 12S core particle to form the 15S particle, followed by incorporation of SF3a to form the mature 17S particle (Fig. 12.2a–b) (Kramer et al. 1999). Although a standalone high resolution structure of the full 17S U2 snRNP has not yet been reported, there is considerable structural data when docked to the U2-snRNA/intron branch helix in the context of the A, pre-B, B, and B^{act} structures. Indeed, as will be discussed shortly, the interactions between the branch helix and SF3b are well resolved in several cryo-EM structures. Nevertheless, determining the standalone U2 snRNP structure will be very important for better understanding the conformational changes required for unwinding the BSL and loading the BPRS onto the pre-mRNA for BS selection. Structural elucidation of this loading process will be a critical aspect for mechanistically understanding certain disease relevant mutations that will be discussed later.

Of the seven SF3b components, perhaps the two most interesting are SF3B1 and PHF5a (Fig. 12.5a, b). SF3B1 is a ~145 kD protein divided between an unstructured N-terminal region followed by 20 HEAT repeats (Fig. 12.5a). The N-terminal domain contains 5 U2AF ligand motifs (ULMs) and a P14 binding sequence, both of which will be discussed in later sections (Fig. 12.5a). More important are the HEAT repeats, each consisting of a pair of alpha helices (about 40 amino acids total) that fold back on themselves to form a zig-zag pattern (Fig. 12.5a). Twenty such repeats can form long extended structures with higher order topologies. Such zig-zagging HEAT repeats, and variations thereof, are fairly typical for scaffolding proteins involved in protein–protein or protein-RNA interactions (Allan and Ratajczak 2011; Yoshimura and Hirano 2016). The first clue about SF3B1 function came from crosslinking studies that suggested it binds to the BP sequence and interacts with U2AF2 (Gozani et al. 1998).

Cryo-EM and crystal structures show that the 20 HEAT repeats pack to form approximately one full turn of a superhelical coil resembling a split lock washer (Fig. 12.5c). Nestled against the inner side where the two ends of the washer come together is the remarkable protein PHF5a (Fig. 12.5b, c). PHF5a is a small globular protein of only ~12 kD. Topologically, it forms an unusual trefoil knot (Fig. 12.5b), which is stabilized by 3 zinc finger motifs (Teng et al. 2017). Knotted folds of any topology are uncommon in the Protein Data Bank, occurring less than 1% of the

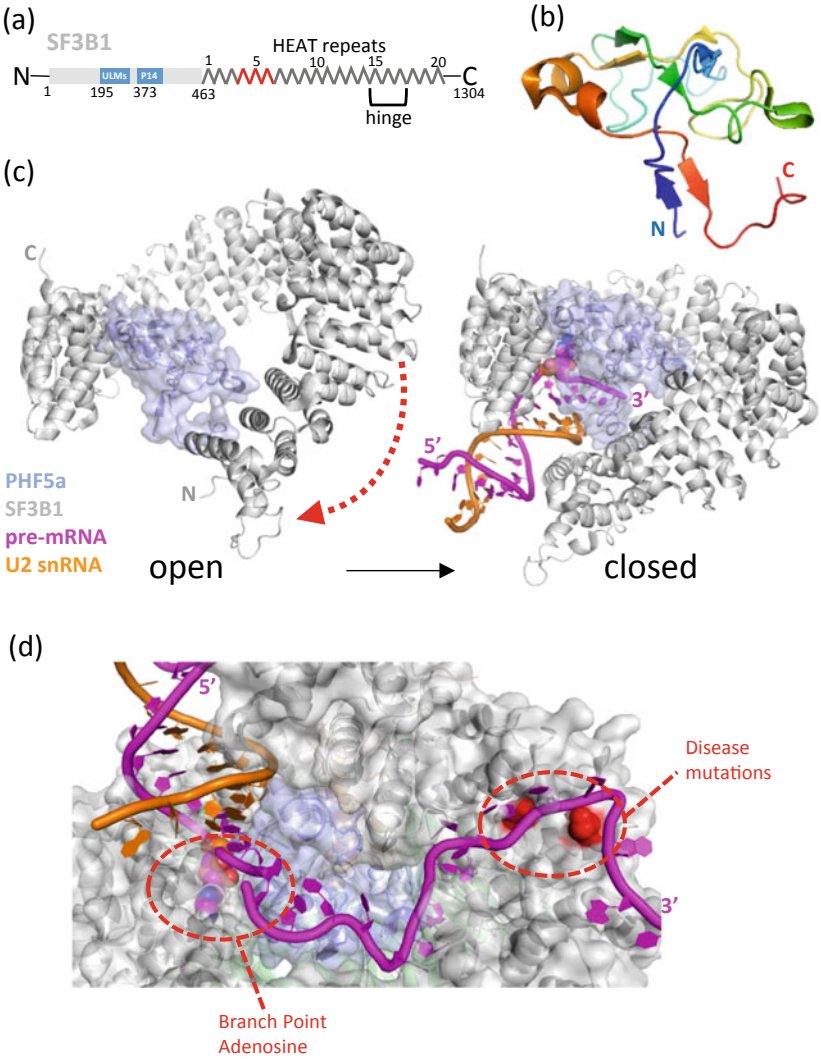


Fig. 12.5 **a** Schematic of SF3B1 protein. HEAT repeats in red correspond to locations of disease mutations. **b** Cartoon rendering of PHF5a to illustrate the unusual knotted topology. Protein colored in rainbow from blue at N terminus to red at C terminus. **c** SF3B1 undergoes dramatic conformational change upon RNA binding. The apo crystal structure and RNA-bound cryo-EM structures are aligned using PHF5a as reference frame. The red arrow indicates how the N-terminal HEAT repeats swing down relative to PHF5a and clamps onto the branch helix like a molecular pincer. The hinge region for this movement occurs in HEAT repeats 15–17. The branch point adenosine is buried in the interface between SF3B1 and PHF5a and rendered in spheres. **d** Model illustrating the position of the most frequently occurring SF3B1 disease mutations K666 and K700E (highlighted in red) in relation to the branch point adenosine and pre-mRNA. This figure was generated from the human B^{act} cryo-EM structure and modeling in the corresponding pre-mRNA (PY tract) from the yeast B^{act} structure where not resolved in the human. These mutations lead to incorrect branch point selection and mis-splicing for certain substrates

time. Primary sequence and 3-D tertiary fold analysis shows that PHF5a has no close homologs in the genome, suggesting a unique biological role for this protein. Indeed, the interface between SF3B1 and PHF5a forms a small tunnel, and the cryo-EM structures of the A, pre-B, B, and B^{act} complexes shows this is where the BP adenosine binds (Fig. 12.2c), validating the aforementioned crosslinking studies (Gozani et al. 1998). As previously discussed, the BP adenosine is singled out through base pairing of neighboring branch site bases with the U2 snRNA (Fig. 12.4c). The BP adenosine is unpaired in this branch helix and in a flipped-out conformation that is held in place by the SF3B1-PHF5a binding pocket through hydrophobic interactions (Fig. 12.3c). Importantly, there are no observed H-bond interactions formed with the base, so that the branch point adenosine selectivity is strictly defined by the pairing interactions with the U2 snRNA (Fig. 12.4c). In the cryo-EM structures of the human spliceosome, the pre-mRNA is not resolved more than ~5 bases after the branch point because the complex was purified using a truncated pre-mRNA substrate that lacked the 3′ SS. In the yeast structures, however, the pre-mRNA is visible as it threads its way (5′–3′) along the bottom face SF3B1. Alignment of the yeast and human structures shows the pre-mRNA would exit SF3B1 over the bottom of HEAT repeats 5–6 in the human structure where the disease mutations most frequently occur (Fig. 12.5c). Just beyond these repeats is approximately where the 3′ SS would be located, although it is not resolved in any of the SF3b containing cryo-EM structures.

Sitting on the top face of SF3B1 and PHF5a is SF3B3 (Fig. 12.6a). SF3B3 consists of three beta-propellers, which are also very typical scaffolding motifs known to mediate protein–protein interactions. The small 10 kD extended helical protein SF3B5 bridges SF3B3, SF3B1, and PHF5a, acting like a staple to hold all four proteins together (Fig. 12.6a). This stand-alone four-protein complex (SF3B1 HEAT domain, PHF5A, SF3B3, and SF3B5) can be recombinantly co-expressed and purified (Teng et al. 2017). The crystal structure shows SF3B1 adopts a relatively extended or open conformation differing substantially from its closed conformation when bound to the RNA-branch helix (Fig. 12.5c) (Cretu et al. 2016). Comparison of the open and closed conformation shows an apparent hinge between HEAT repeats 15–17, which causes the later repeats to swing down and clamp onto the branch helix like a molecular pincer (Fig. 12.5a, c) (Cretu et al. 2018; Finci et al. 2018). The open conformation of the four-protein subcomplex binds the three known classes of natural product inhibitors (Cretu et al. 2018) and will be discussed further in a subsequent section. Interestingly, while the four-protein complex binds inhibitors, we have been unable to demonstrate clear evidence of binding to ~50 to 100-mer synthetic oligos of single stranded or double stranded duplex RNA, which suggests other components are necessary for loading and/or cooperatively stabilizing SF3b onto RNA. Interestingly, SF3b can be immunodepleted from nuclear extracts, and recombinant seven-protein SF3b complex can be added back in to reconstitute splicing activity, implicating other members of the SF3b complex as important accessory factors for cooperative integration with the U2 snRNA and the branch helix (Cretu et al. 2016).

In addition to the 4 proteins discussed above (SF3B1, PHF5a, SF3B3, and SF3B5), the SF3b complex also contains SF3B2, SF3B4, and P14 (Fig. 12.6a–c). The cryo-EM structures show that the partially resolved SF3B2 forms extensive interactions with

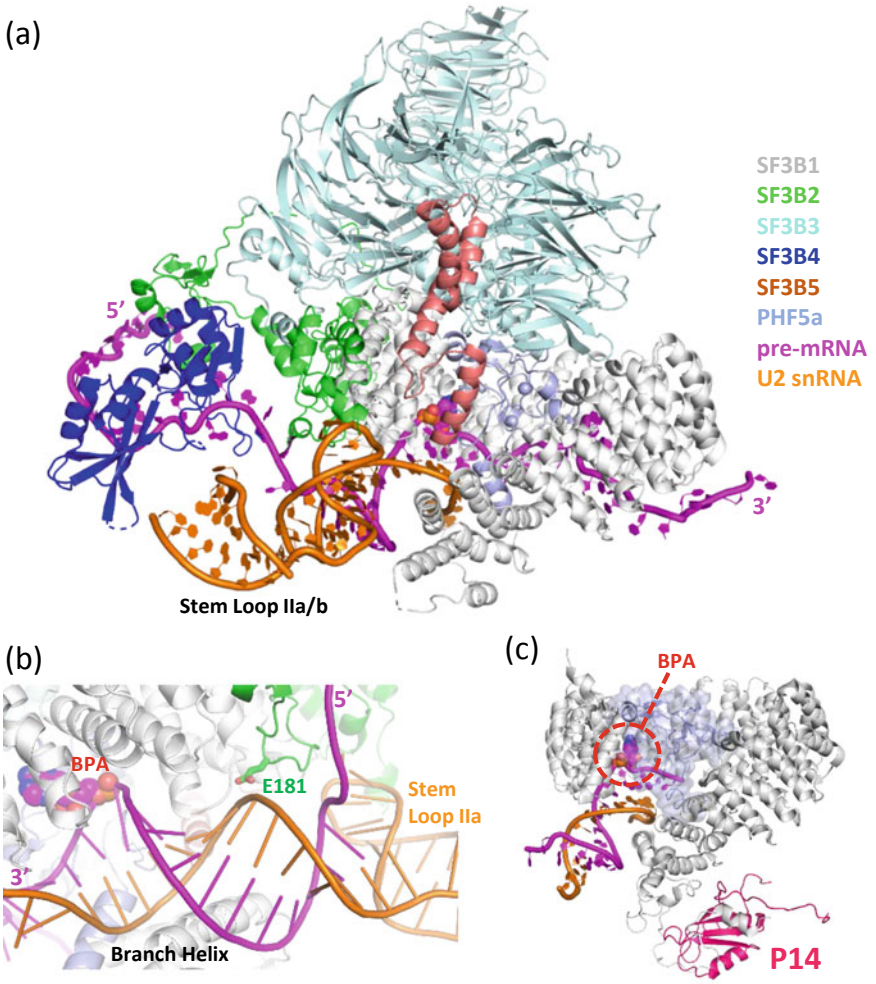


Fig. 12.6 Overview of the full SF3b structure. **a** Figure generated from the yeast B complex. SF3B3 sits on the top face of SF3B1 and PHF5a. SF3B5 forms an extended helical structure that stiches the four proteins together. SF3B2 forms contacts on the outer face of SF3B1, SF3B4 and both the pre-mRNA and U2 snRNA. SF3B4 also interacts with both the pre-mRNA and snRNA. **b** The protein Cus2 (SF3B2 in humans) has a known mutation E181K that rescues mutations in Stem Loop IIa. K181 would enhance electrostatic interactions with the RNA backbone. **c** The protein P14 does not exist in yeast. In humans, it has been resolved in the human B^{act} cryo-EM structure and on the outer face of SF3B1 near HEAT repeat 1–2 and about 50 Å from the branch point adenosine (BPA)

SF3B4, the outer face of the SF3B1 HEAT repeats, the RNA branch duplex, and Prp9 from the SF3a complex. SF3B2 (CUS1 in yeast) was first characterized functionally in yeast as one of two proteins that were suppressors of cold sensitive U2 snRNA (Cus1 and Cus2) mutant growth phenotype (Wells et al. 1996; Pauling et al. 2000). These U2 snRNA mutant strains have a weakened base pair interaction in stem loop IIa that inhibits growth at low temperatures. Gratifyingly, the cryo-EM structure (Yan et al. 2016) illustrates how the suppressor mutation CUS1-E181K would enhance favorable electrostatic interactions with the U2 snRNA phosphate backbone in the branch helix and therefore have a stabilizing effect that would compensate for folding defects in the adjacent stem loop IIa and therefore rescue growth (Fig. 12.6b).

The interactions between SF3B2 and SF3B4 have been resolved at higher resolution by X-ray crystallography (van Roon et al. 2017), and these interactions are mirrored in the cryo-EM structure. The full length SF3B4 is not fully resolved by cryo-EM, but its RRM domains are visible, with the first RRM interacting with SF3B2 and the pre-mRNA upstream of the branch sequence, and the second RRM interacting with PRP9 and the U2snRNA stem loop II (Fig. 12.6a). The last protein of the SF3b complex to be discussed is the small and globular 14 kD protein P14 (Fig. 12.6c). Its precise role in splicing is less clear. To date, P14 has only been resolved in the human B^{act} cryo-EM structure and is located on the external side of SF3B1 near HEAT repeats 1–2 (Haselbach et al. 2018). Early cross-linking and co-IP studies implicated P14 in BP adenosine recognition (Query et al. 1996; Will et al. 2001), but the cryo-EM structures contradict those data with PHF5a forming the main interactions with the BP adenosine as early as the "A complex" (Plaschka et al. 2018) and P14 being well-removed from the BP adenosine (Haselbach et al. 2018). While still possible that P14 could contact the BP adenosine very early during spliceosome assembly, it is also possible that early biochemical studies faced certain technical challenges that may have confounded P14 with PHF5a. Indeed, both proteins have nearly identical molecular weights and, being part of the same complex, co-IP together. Additionally, perhaps because of the knotted topology of PHF5a, the coverage of PHF5a peptides in proteomic mass spectrometry experiments (in our hands) is much lower than for P14. The similarity in molecular weight, the co-association, and the apparent technical challenges of the mass spec could all potentially explain how P14 was singled out. Regardless, it is clear that PHF5a and SF3B1 together form the true BP adenosine-binding pocket, and further studies are warranted to better define the role of P14 in splicing. Interestingly, Saccharomyces cerevisiae lacks a P14 homolog, suggesting an important role as a splicing factor for higher organisms that is dispensable for lower.

Finally, the three protein SF3a complex forms a bridge between the SF3b complex and the core U2 snRNP proteins U2A′, U2B″, and the SM ring (Fig. 12.2a). The SF3a60 (PRP9) protein binds to the SM core proteins and U2 snRNA stem loop II (Lin and Xu 2012). SF3a66 (PRP11) interacts with SF3B1, SF3B2, and the branch helix RNA. SF3a120 (Prp21) is mostly unresolved in the available cryo-EM structures, but a portion of it is visible linking SF3a60 to SF3a66. The SF3a-mediated bridging interactions between the core (SM and U2A′/B″) and SF3b suggest a key role for stabilizing the mature 17 S U2 snRNP.

These SF3a and SF3b sub-complexes, therefore, constitute the dynamic core subunits of the U2 snRNP. However, there are some additional proteins that interact transiently as well and potentially play important roles in the early recruitment of the U2 snRNP to the 3′ SS and in alternative splicing. SF3B1 is again the key player in mediating these protein–protein interactions. As previously mentioned, the N-terminal domain of SF3B1 is unstructured but contains 5 U2AF ligand motifs (ULMs) that are known to bind with U2AF homology motif (UHM) containing proteins (Fig. 12.5a). Unfortunately, these ULM motifs have not yet been resolved in the cryo-EM structures. Nevertheless, several of these partners have been characterized to date, including U2AF2 (Thickman et al. 2006), which may be important for recruitment to the 3′ SS, SPF45 (Corsini et al. 2007), RBM39 (Loerch et al. 2014) and TAT-SF1 (Loerch et al. 2019). Indeed, there are crystal structures of the ULM5 motif of SF3B1 bound to three of these splicing factors including SPF45, RBM39, and TAT-SF1 (Corsini et al. 2007; Stepanyuk et al. 2016; Loerch et al. 2019). Of these, TAT-SF1 is noteworthy because the corresponding protein homolog in yeast (Cus2) was also identified as a suppressor of the cold sensitive U2 snRNA mutant growth phenotype and is thought to help in the folding of U2 snRNA to facilitate recruitment and binding of SF3b and SF3a (Wells et al. 1996; Yan et al. 1998). TAT-SF1 co-immunoprecipitates with the U2 snRNP (Agafonov et al. 2011) but has not yet been visualized by cryo-EM in the available later stage spliceosome structures (consistent with an earlier role in U2 snRNP assembly). It will be extremely interesting to see how these splicing factors interact with RNA and other protein subunits of the U2 snRNP and may provide additional insights into the structural basis for alternative splicing.

SF3B1 Disease Mutations

SF3B1 is one of several splicing proteins that is frequently mutated in both hematologic cancers and solid tumors (Yoshida and Ogawa 2014; Agrawal et al. 2018). Recurring hotspot mutations are found in myelodysplastic syndromes (MDS) (Yoshida et al. 2011), chronic lymphocytic leukemia (CLL) (Landau et al. 2013), and chronic myelomonocytic leukemia (CMML) (Patnaik et al. 2013), among others. These mutations are mutually exclusive, heterozygous, and lead to mis-splicing to form aberrant transcripts. Many of these transcripts are targeted for nonsense-mediated decay (NMD), leading to their downregulation (Darman et al. 2015), but a subset could also be translated with potentially neomorphic activities that contribute to disease. For example, in uveal melanoma, SF3B1 mutation leads to mis-splicing of BRD9, which compromises its normal function as a tumor suppressor. Correcting this mis-splicing event using CRISPR technology suppresses tumor growth, singling out this neomorphic activity as a possible oncogenic driver event (Inoue et al. 2019). Thus, mutations in SF3B1 and other splicing factors have been shown to lead to mis-splicing and a “spliceosome sickness” phenotype where cells are dependent on their remaining wild type copy for survival (Darman et al. 2015).

The most frequent mutations in SF3B1 include the K666N and the K700E mutations (Fig. 12.5c). Extensive biochemical characterization of these mutants using in vitro splicing assays has demonstrated that these mutants can select, depending on the pre-mRNA substrate, alternative branch point sequences (Darman et al. 2015; Alsafadi et al. 2016). Selection of the wrong BP adenosine can then lead to incorrect exon ligation if there also happens to be a coincident YAG sequence nearby (within 10–15 bases immediately downstream) (Horowitz 2012). This noncanonical YAG has been termed the cryptic AG or cryptic 3′ SS since it may appear upstream of the PY tract. The result of this mis-splicing is the incorporation of non-coding sequence into the transcript and potentially premature termination.

Interestingly, these SF3B1 mutations are located on the bottom face of HEAT repeats 5–6, precisely where the 3′ end of the pre-mRNA is expected to be located (Fig. 12.5a, c). This has led to speculation that the change from basic to neutral or acidic residue could adversely affect electrostatic interactions between SF3B1 and the phosphate backbone of the mRNA, which, consequently, could result in loss of fidelity for selection of the BP adenosine. However, there are obvious limitations to this model. Recent work has examined this mechanistic question more closely using yeast as a model system. Here, the corresponding disease mutations were engineered into the yeast sequence and were shown to phenocopy mis-splicing observed in the human system. Moreover, it was shown that interactions between SF3B1 and the helicase Prp5 were compromised (Tang et al. 2016). As previously discussed, the helicase Prp5 is necessary for unwinding the branch stem loop (BSL), and pairing the branch point recognition sequence (BPRS) of the U2 snRNA with the branch site sequence in the pre-mRNA, and loading the resulting branch helix onto the SF3b complex. This ATP-dependent step is required for recruiting the U2 snRNP to the 3′ SS. Therefore, the mutations in SF3B1 likely weaken the interaction with the helicase Prp5, leading to loss in fidelity for U2 snRNA pairing and selection of the incorrect branch point adenosine (Tang et al. 2016). Future structural studies with WT and mutant SF3B1 with Prp5 will be required to visualize and understand this process for selection of the branch sequence and loading of the branch helix onto SF3b.

Finding SF3B1 mutant selective small molecules could be an attractive and useful therapeutic strategy for cancers harboring these mutations. At least two possible approaches could be considered. One approach would be identifying a selective compound that would inhibit mutant SF3b activity but spare WT. A second possibility could be identifying a compound that rescues the compromised mutant activity. Either way, both strategies face daunting challenges. The SF3B1 mutations are located at the bottom of the HEAT repeats on a flat interface with no obvious binding pocket, and difficulties in reconstituting biochemical activity from minimal complexes in vitro limit compound library screening opportunities. Moreover, mutant phenotypic screens have not yielded satisfactory hits. Therefore, finding mutant selective compounds to either inhibit or rescue mutant activity is an extremely challenging area for future drug discovery.

Splicing Modulators

While the prospect of finding SF3B1 mutant selective inhibitors remains a challenge for drug discovery, progress has been made in the discovery of non-selective modulators. These could have utility in a clinical setting. Cells harboring mutations in splicing components are heterozygous and dependent on their remaining WT copy for viability (Zhou et al. 2015). This haploinsufficiency might give a therapeutic window that can be exploited (Obeng et al. 2016; Paolella et al. 2017; Seiler et al. 2018). Their utility is also being explored pre-clinically in the context of cytotoxic agents specifically targeted using antibody drug conjugates (ADCs) (Puthenveetil et al. 2016). This is a significant area for research because alternative ADC modalities will be needed in the clinic to help overcome emerging resistance to existing treatment options (Garcia-Alonso et al. 2020).

The first reported splicing inhibitors were discovered serendipitously from phenotypic screens (Miller-Wideman et al. 1992; Nakajima et al. 1996; Mizui et al. 2004). The best characterized of these are natural products derived from three different classes: herboxidiene, spliceostatin, and pladienolides (Fig. 12.7). Extensive chemical biology and target identification efforts determined that all three classes of

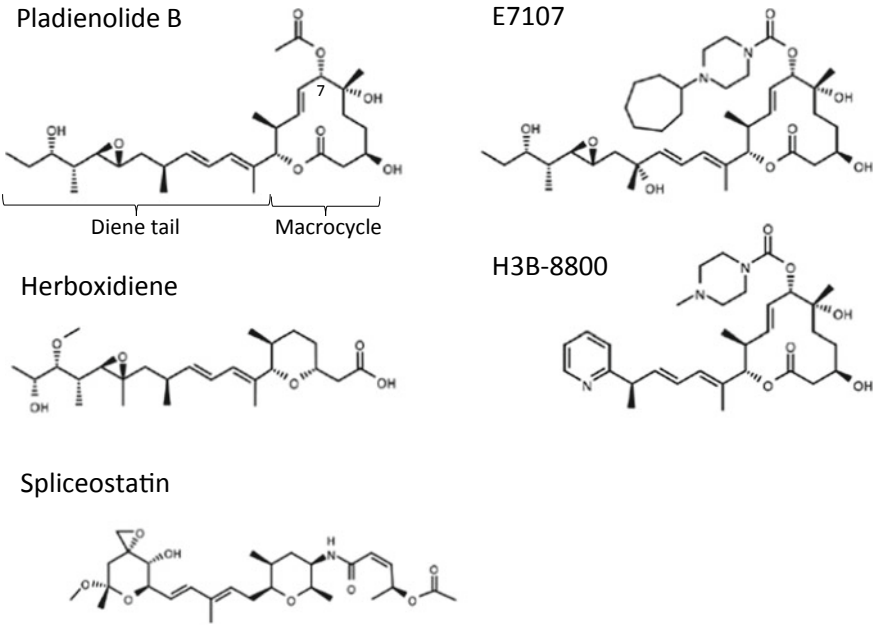


Fig. 12.7 Natural product and natural product based inhibitors described in the text. The diene tail is aligned for Plad B and herboxidiene and these two compounds likely will share similar binding modes. The natural product spliceostatin has a distinct structure and will likely have a different binding mode. H3B-8800 has a shorter diene tail which likely explains its weaker activity in viability and in vitro splicing assays compared to E7107

compound interact with SF3B1 (Kaida et al. 2007; Kotake et al. 2007; Hasegawa et al. 2011). Further chemogenomic work identified drug resistance mutations in SF3B1 (Yokoi et al. 2011) and PHF5a (Teng et al. 2017). Analysis of the B^{act} cryo-EM structure showed that these resistance mutations cluster within the BP adenosine-binding pocket (Teng et al. 2017). The two mutations that impart the greatest resistance are SF3B1 R1074H and PHF5a Y36C mutation. Incredibly, while these two residues form contacts with the branch point adenosine, overexpression of these mutants has no effect on viability and shows strong resistance to compound (Teng et al. 2017). Crystal structures and cryo-EM structures now confirm that these compounds occupy the BP adenosine pocket interacting with both SF3B1 and PHF5a and form direct contacts with SF3B1 R1074 and PHF5a Y36 (Cretu et al. 2018; Finci et al. 2018). Of these three classes of natural product inhibitors, the furthest to advance into the clinic are E7107 (Hong et al. 2014) and H3B-8800 (Buonamici et al. 2016) (Fig. 12.7), both of which derive from the pladienolide scaffold and were separately optimized extensively for potency and favorable PK, respectively. The highest resolution structure available is with the related compound pladienolide B (Plad B) bound to the 4-protein complex SF3B1-PHF5a-SF3B3-SF3B5 (Cretu et al. 2018).

The structure with Plad B shows how the diene tail projects into the pocket where the branch point adenosine would otherwise bind and packs against Y36 of PHF5a, which was one of the identified resistance mutations (Fig. 12.8a, b). Compounds with truncated diene tails do not reach far enough into the pocket and therefore have no appreciable inhibitory activity (Finci et al. 2018). Herboxidiene, which contains an almost identical diene tail to Plad B (Fig. 12.7), is expected to have a similar binding mode. Additionally, both the available crystal structure with Plad B and the lower resolution cryo-EM structure with E7107 show that SF3B1 is in the same conformation as the RNA-unbound state. This revelation immediately suggests that these inhibitors, in addition to interfering directly with branch point adenosine binding, may also act as wedges that hold SF3B1 in an open conformation incompatible with RNA branch helix binding (Fig. 12.8b) (Cretu et al. 2018; Finci et al. 2018). The third class of compounds, spliceostatin and its derivatives, has a very different chemical structure from herboxidiene or pladienolide. It is anticipated that this class of compounds will bind in the same pocket since they, too, are susceptible to the same resistance mutations, but their binding mode is expected to be quite different. Therefore, how this distinct scaffold binds remains an open question.

Previous work has shown that for certain pre-mRNA substrates, modulators appear to stall the spliceosome in the "A-complex" (Roybal and Jurica 2010) and may weaken the interaction between the U2 snRNP and the pre-mRNA (Roybal and Jurica 2010; Folco et al. 2011). Consistent with this observation, for inhibition to occur, compounds must be added to nuclear lysates before initiating in vitro splicing reactions with pre-mRNA substrate and ATP. Splicing modulation does not occur if compounds are added after the pre-mRNA substrate and ATP (Folco et al. 2011). Together, these observations suggest that the modulators impinge on the splicing reaction early, probably affecting the formation of the "A-complex". These data are consistent with the structures, which show that SF3B1 is in the closed conformation when bound to the branch helix as early as the "A complex". In other words, these

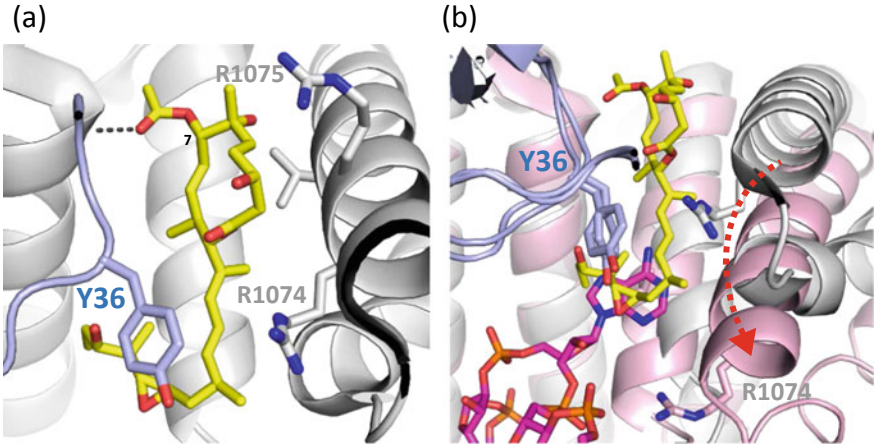


Fig. 12.8 Binding pose for Pladienolide B. **a** Crystal structure showing the binding pose of Pladienolide B in the branch point adenosine binding pocket. SF3B1 is rendered in gray and PHF5a in blue. The key resistance mutations Y36 and R1074 form key interactions with the ligand. The acetyl group at C7 forms an H-bond interaction with the backbone amide of R38. **b** A zoomed out view showing the overlay of the Plad B bound structure (SF3B1 in gray) with the corresponding RNA bound structure from the human Bact complex (SF3B1 in pink and pre-mRNA in magenta). The overlay is generated using PHF5a as the reference frame. The overlay illustrates the steric clash that would occur between the diene tail and the adenosine and the substantial movement in the R1074 containing helix, indicated by the red arrow. Inhibitors wedge SF3B1 in this open conformation. In the RNA bound structure, Y36 packs against the adenine ring and R1074 provides charge complementarity for the pre-mRNA phosphate backbone

compounds would be less likely to be effective inhibitors once the "A complex" has assembled. Therefore, as a corollary, these compounds could serve as excellent chemical probes to isolate WT and mutant SF3b containing U2 snRNP complexes for future structural analysis.

Global RNAseq analysis shows that none of these compounds is a universal inhibitor across all pre-mRNA substrates. Even at doses tenfold greater than their GI50 for a particular cell line, these compounds impact at most 10% of all possible splicing junctions. These inhibited junctions usually manifest as intron-inclusion events as well as some exon-skipping events (Corrionero et al. 2011; Buonamici et al. 2016). Indeed, H3B-8800, which is less potent in viability assays and in vitro splicing assays, affects far fewer splicing junctions than a more potent compound like E7107 (Seiler et al. 2018). Given the shorter diene tail of H3B-8800, it is not expected to project as far into the branch point adenosine pocket, and that would explain its weaker activity. RNAseq data shows that pre-mRNA substrates more likely to be inhibited by these modulators are shorter in length and higher in relative GC content (Buonamici et al. 2016; Teng et al. 2017). The degree to which this interaction with the branch point is inhibited depends, obviously, on the dose of compound but also on the relative strength of the pre-mRNA substrate (Finci et al. 2018), which explains why not all pre-mRNA introns are universally inhibited when

analyzed in global RNAseq (Buonamici et al. 2016). While these rules for inhibition have not been fully defined, contributing factors may include the degree of complementarity between the pre-mRNA and U2 snRNA, relative GC content, intron length, and likely a host of other cooperative interactions with additional splicing factors (Vigevani et al. 2017). Regardless, these data are all consistent with a pre-mRNA substrate competitive mechanism of action for these compounds, where weaker pre-mRNA substrates are unable to effectively compete with the inhibitor, which leads to intron retention or exon skipping. In contrast, stronger pre-mRNA substrates are able to effectively compete with the inhibitor and, therefore, splicing can proceed normally.

Summary

In this chapter, we have provided a broad overview of splicing and the new structural insights gleaned from cryo-EM structures determined over the past few years. The rich information content of these data has provided an overarching context for interpreting and, in some cases, reinterpreting three decades' worth of extraordinary yeast genetics and spliceosome biochemistry. New binding partners are now revealed and appreciated, which will help inform more focused analysis to better understand the mechanistic relationships between different proteins. We have also reviewed the composition and various structural features of the U2 snRNP and how it fits into the broader context of the spliceosome. The SF3b complex plays a pivotal role in splicing for sequestering the branch point adenosine during the early assembly phase of the spliceosome. SF3B1 and PHF5a are the two key proteins that wrap around the branch point adenosine. Frequently occurring disease mutations are located in SF3B1 and are known to compromise the fidelity of branch point selection. Recent functional characterization of these mutations in yeast points to compromised interactions with Prp5 as the possible explanation. A better understanding of branch helix loading and the ability to reconstitute this in vitro using recombinant and purified components will open new doors and new screening opportunities for the ambitious drug discovery goal of finding mutant selective compounds that either inhibit or rescue mutant activity. Finally, we reviewed the functions of SF3B1 and PHF5A in sequestering the branch point adenosine. Remarkably, three completely different classes of natural product-based inhibitors mechanistically converge to impinge this single pocket in the spliceosome. Incredible progress has been made in our understanding of these compounds as pre-mRNA substrate competitive inhibitors. Future efforts and investment will be needed to bring spliceosome-based therapies and treatment options to patients.

References

Agafonov DE, Deckert J et al (2011) Semiquantitative proteomic analysis of the human spliceosome *via* a novel two-dimensional gel electrophoresis method. Mol Cell Biol 31(13):2667–2682
Agrawal AA, Salsi E et al (2016) An extended U2AF(65)-RNA-binding domain recognizes the 3' splice site signal. Nat Commun 7:10950
Agrawal AA, Yu L et al (2018) Targeting splicing abnormalities in cancer. Curr Opin Genet Dev 48:67–74
Allan RK, Ratajczak T (2011) Versatile TPR domains accommodate different modes of target protein recognition and function. Cell Stress Chaperones 16(4):353–367
Alsafadi S, Houy A et al (2016) Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat Commun 7:10615
Bai R, Wan R et al (2018) Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. Science 360(6396):1423–1429
Bai R, Yan C, et al (2017) Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. Cell 171(7):1589–1598 e1588
Bertram K, Agafonov DE, et al (2017a) Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. Cell 170(4):701–713 e711
Bertram K, Agafonov DE, et al (2017) Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature 542(7641):318–323
Brosi R, Hauri HP et al (1993) Separation of splicing factor SF3 into two components and purification of SF3a activity. J Biol Chem 268(23):17640–17646
Buonamici S, Yoshimi A et al (2016) H3B-8800, an orally bioavailable modulator of the SF3b complex, shows efficacy in spliceosome-mutant myeloid malignancies. Blood 128:966
Charenton C, Wilkinson ME et al (2019) Mechanism of 5' splice site transfer for human spliceosome activation. Science 364(6438):362–367
Chen HC, Tseng CK et al (2013) Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. Mol Cell Biol 33(3):514–525
Corrionero A, Minana B et al (2011) Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev 25(5):445–459
Corsini L, Bonnal S et al (2007) U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat Struct Mol Biol 14(7):620–629
Cretu C, Agrawal AA, et al (2018) Structural basis of splicing modulation by antitumor macrolide compounds. Mol cell 70(2):265–273 e268
Cretu C, Schmitzova J et al (2016) Molecular architecture of SF3b and structural consequences of its cancer-related mutations. Mol Cell 64(2):307–319
Darman RB, Seiler M et al (2015) Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. Cell Rep 13(5):1033–1045
Fica SM, Oubridge C et al (2017) Structure of a spliceosome remodelled for exon ligation. Nature 542(7641):377–380
Fica SM, Oubridge C et al (2019) A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. Science 363(6428):710–714
Finci LI, Zhang X et al (2018) The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. Genes Dev 32(3–4):309–320
Folco EG, Coil KE et al (2011) The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. Genes Dev 25(5):440–444
Galej WP, Wilkinson ME et al (2016) Cryo-EM structure of the spliceosome immediately after branching. Nature 537(7619):197–201
Garcia-Alonso S, Ocana A et al (2020) Trastuzumab emtansine: mechanisms of action and resistance, clinical progress, and beyond. Trend Cancer 6(2):130–146
Golas MM, Sander B et al (2003) Molecular architecture of the multiprotein splicing factor SF3b. Science 300(5621):980–984

Gozani O, Potashkin J et al (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol Cell Biol 18(8):4752–4760
Hang J, Wan R et al (2015) Structural basis of pre-mRNA splicing. Science 349(6253):1191–1198
Hasegawa M, Miura T et al (2011) Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. ACS Chem Biol 6(3):229–233
Haselbach D, Komarov I et al (2018) Structure and conformational dynamics of the human spliceosomal B(act) complex. Cell 172(3):454–464 e411
Hong DS, Kurzrock R et al (2014) A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicesome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. Invest New Drugs 32(3):436–444
Horowitz DS (2012) The mechanism of the second step of pre-mRNA splicing. Wiley Interdisc Rev RNA 3(3):331–350
Inoue D, Chew GL et al (2019) Spliceosomal disruption of the non-canonical BAF complex in cancer. Nature 574(7778):432–436
Kaida D, Motoyoshi H et al (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol 3(9):576–583
Kastner B, Will CL et al (2019) Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. Cold Spring Harbor Persp Biol 11(11)
Kondo Y, Oubridge C, et al (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. eLife 4
Kotake Y, Sagane K et al (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol 3(9):570–575
Kramer A, Gruter P et al (1999) Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J Cell Biol 145(7):1355–1368
Landau DA, Carter SL et al (2013) Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152(4):714–726
Lardelli RM, Thompson JX et al (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA 16(3):516–528
Li X, Liu S et al (2019) A unified mechanism for intron and exon definition and back-splicing. Nature 573(7774):375–380
Lin PC, Xu RM (2012) Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J 31(6):1579–1590
Liu S, Li X et al (2017) Structure of the yeast spliceosomal postcatalytic P complex. Science 358(6368):1278–1283
Liu Z, Luyten I et al (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. Science 294(5544):1098–1102
Loerch S, Leach JR et al (2019) The pre-mRNA splicing and transcription factor Tat-SF1 is a functional partner of the spliceosome SF3b1 subunit *via* a U2AF homology motif interface. J Biol Chem 294(8):2892–2902
Loerch S, Maucuer A et al (2014) Cancer-relevant splicing factor CAPERalpha engages the essential splicing factor SF3b155 in a specific ternary complex. J Biol Chem 289(25):17325–17337
Miller-Wideman M, Makkar N et al (1992) Herboxidiene, a new herbicidal substance from Streptomyces chromofuscus A7847. Taxonomy, fermentation, isolation, physico-chemical and biological properties. J Antibio 45(6):914–921
Mizui Y, Sakai T et al (2004) Pladienolides, new substances from culture of *Streptomyces platensis* Mer-11107. III. In vitro and in vivo antitumor activities. J Antibio 57(3):188–196
Nakajima H, Sato B et al (1996) New antitumor substances, FR901463, FR901464 and FR901465. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activities. J Antibio 49(12):1196–1203
Newman AJ, Norman C (1992) U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell 68(4):743–754

Obeng EA, Chappell RJ et al (2016) Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. Cancer Cell 30(3):404–417

Ohrt T, Odenwalder P et al (2013) Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA 19(7):902–915

Paolella BR, Gibson WJ, et al (2017) Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. eLife 6

Patnaik MM, Lasho TL et al (2013) Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. Am J Hematol 88(3):201–206

Pauling MH, McPheeters DS et al (2000) Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. Mol Cell Biol 20(6):2176–2185

Perriman R, Ares M Jr (2010) Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. Mol Cell 38(3):416–427

Plaschka C, Lin PC et al (2017) Structure of a pre-catalytic spliceosome. Nature 546(7660):617–621

Plaschka C, Lin PC et al (2018) Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature 559(7714):419–422

Pomeranz Krummel DA, Oubridge C et al (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature 458(7237):475–480

Puthenveetil S, Loganzo F et al (2016) Natural product splicing inhibitors: a new class of antibody-drug conjugate (ADC) payloads. Bioconjug Chem 27(8):1880–1888

Query CC, Strobel SA et al (1996) Three recognition events at the branch-site adenine. EMBO J 15(6):1392–1402

Rauhut R, Fabrizio P et al (2016) Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. Science 353(6306):1399–1405

Roybal GA, Jurica MS (2010) Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. Nucleic Acids Res 38(19):6664–6672

Seiler M, Yoshimi A et al (2018) H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat Med 24(4):497–504

Shi Y (2017) The spliceosome: a protein-directed metalloribozyme. J Mol Biol 429:2640–2653

Sontheimer EJ, Steitz JA (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science 262(5142):1989–1996

Stepanyuk GA, Serrano P, et al (2016) UHM-ULM interactions in the RBM39-U2AF65 splicing-factor complex. Acta Crystallograph Sect D Struct Biol 72(Pt 4):497–511

Tang Q, Rodriguez-Santiago S et al (2016) SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes Dev 30(24):2710–2723

Teng T, Tsai JH et al (2017) Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. Nat Commun 8:15522

Thickman KR, Swenson MC et al (2006) Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. J Mol Biol 356(3):664–683

Toroney R, Nielsen KH et al (2019) Termination of pre-mRNA splicing requires that the ATPase and RNA unwindase Prp43p acts on the catalytic snRNA U6. Genes Dev 33(21–22):1555–1574

Tseng CK, Liu HL et al (2011) DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA 17(1):145–154

van der Feltz C, Hoskins AA (2019) Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. Crit Rev Biochem Mol Biol 54(5):443–465

van Roon AM, Oubridge C et al (2017) Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p-binding domain. RNA 23(6):968–981

Vigevani L, Gohr A et al (2017) Molecular basis of differential 3' splice site sensitivity to anti-tumor drugs targeting U2 snRNP. Nat Commun 8(1):2100

Wan R, Bai R, et al (2019a) Structures of the catalytically activated yeast spliceosome reveal the mechanism of branching. Cell 177(2):339–351 e313
Wan R, Bai R, et al (2019b) How is precursor messenger RNA spliced by the spliceosome? Ann Rev Biochem
Wang W, Maucuer A et al (2013) Structure of phosphorylated SF1 bound to U2AF(6)(5) in an essential splicing factor complex. Structure 21(2):197–208
Wan R, Yan C et al (2016) Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. Science 353(6302):895–904
Wan R, Yan C, et al (2017) Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae.* Cell 171(1):120–132 e112
Wells SE, Neville M et al (1996) CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. Genes Dev 10(2):220–232
Wilkinson ME, Charenton C, et al (2019) RNA splicing by the spliceosome. Ann Rev Biochem
Wilkinson ME, Fica SM et al (2017) Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. Science 358(6368):1283–1288
Will CL, Schneider C et al (2001) A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. EMBO J 20(16):4536–4546
Xu YZ, Newnham CM et al (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. EMBO J 23(2):376–385
Yan C, Hang J et al (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 349(6253):1182–1191
Yan D, Perriman R et al (1998) CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. Mol Cell Biol 18(9):5000–5009
Yan C, Wan R et al (2016) Structure of a yeast activated spliceosome at 3.5 A resolution. Science 353(6302):904–911
Yan C, Wan R et al (2017) Structure of a yeast step II catalytically activated spliceosome. Science 355(6321):149–155
Yokoi A, Kotake Y et al (2011) Biological validation that SF3b is a target of the antitumor macrolide pladienolide. FEBS J 278(24):4870–4880
Yoshida K, Ogawa S (2014) Splicing factor mutations and cancer. Wiley Interdisc Rev RNA 5(4):445–459
Yoshida K, Sanada M et al (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478(7367):64–69
Yoshimura SH, Hirano T (2016) HEAT repeats—versatile arrays of amphiphilic helices working in crowded environments? J Cell Sci 129(21):3963–3970
Zhang X, Yan C, et al (2017) An atomic structure of the human spliceosome. Cell 169(5):918–929 e914
Zhang X, Yan C et al (2018) Structure of the human activated spliceosome in three conformational states. Cell Res 28(3):307–322
Zhan X, Yan C et al (2018) Structure of a human catalytic step I spliceosome. Science 359(6375):537–545
Zhou Q, Derti A et al (2015) A chemical genetics approach for the functional assessment of novel cancer genes. Can Res 75(10):1949–1958

Chapter 13
Interaction Networks of Ribosomal Expansion Segments in Kinetoplastids

Quentin Vicens, Anthony Bochler, Amy Jobe, Joachim Frank, and Yaser Hashem

Abstract Expansion segments (ES) are insertions of a few to hundreds of nucleotides at discrete locations on eukaryotic ribosomal RNA (rRNA) chains. Some cluster around 'hot spots' involved in translation regulation and some may participate in biogenesis. Whether ES play the same roles in different organisms is currently unclear, especially since their size may vary dramatically from one species to another and very little is known about their functions. Most likely, ES variation is linked to adaptation to a particular environment. In this chapter, we compare the interaction networks of ES from four kinetoplastid parasites, which have evolved in diverse insect vectors and mammalian hosts: *Trypanosoma cruzi, Trypanosoma brucei*, *Leishmania donovani* and *Leishmania major.* Here, we comparatively analyze ribosome structures from these representative kinetoplastids and ascertain meaningful structural differences from mammalian ribosomes. We base our analysis on sequence alignments and three-dimensional structures of 80S ribosomes solved by cryo-electron microscopy (cryo-EM). Striking differences in size are observed between ribosomes of different parasites, indicating that not all ES are expanded equally. Larger ES are not always matched by large surrounding ES or proteins extensions in their vicinity,

Q. Vicens · A. Bochler · Y. Hashem (✉)
Université de Strasbourg, CNRS, Architecture et Réactivité de L'ARN, UPR 9002, 15 rue René Descartes, 67000 Strasbourg, France
e-mail: yaser.hashem@inserm.fr

J. Frank (✉)
Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA
e-mail: jf2192@cumc.columbia.edu

A. Bochler · Y. Hashem
Inserm, Université de Bordeaux, Institut Européen de Chimie et Biologie, U1212, 33600 Pessac, France

Q. Vicens
Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA

A. Jobe · J. Frank
Department of Biological Sciences, Columbia University, New York, NY 10032, USA

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_13

a particularity that may lead to clues about their biological function. ES display different species-specific patterns of conservation, which underscore the density of their interaction network at the surface of the ribosome. Making sense of the conservation patterns of ES is part of a global effort to lay the basis for functional studies aimed at discovering unique kinetoplastid-specific sites suitable for therapeutic applications against these human and often animal pathogens.

Keywords Expansion segment · Kinetoplastid parasite · Ribosome structure · Ribosomal RNA · Eukaryotic translation

Why Expansion Segments?

Because protein synthesis is a universal necessity for all organisms, ribosome function, overall shape and core architecture are conserved across species. Yet, ribosome composition varies considerably among different domains of life. For example, ribosomes from yeast and human are respectively ~40 and ~85% larger than bacterial ribosomes (Melnikov et al. 2012). In addition to comprising more proteins than bacterial ribosomes, eukaryotic ribosomes also contain elongated ribosomal RNAs (rRNA). This 'extra RNA' forms expansion segments (ES) (Clark et al. 1984; Hassouna et al. 1984; Ware et al. 1983).

Whether ES are all mere additions that are tolerated by the translation machinery, or whether they have important regulatory roles remains for the most part an open question. Most ES protrude outside of the highly conserved ribosomal core (Bernier et al. 2018; Melnikov et al. 2012). Their size varies across species, although their patterns of insertion are conserved (Gerbi 1996). We know from early structural studies that ES and associated proteins portions comprise somewhat flexible substructures on the solvent-facing side, where they may interact with one another and with proteins to form an outer layer (Anger et al. 2013; Armache et al. 2010; Gao et al. 2005). This web-like organization is more rigid where contacts with the ribosome core are tighter, which argues against a superfluous nature. ES are generally best perceived as peripheral elements that facilitate folding, assembly, or activity. In this view they are to some extent similar to often observed insertions of >200 nt reported in other ribozymes (e.g., group I introns, RNase P, and hammerhead) (Kachouri et al. 2005; Lehnert et al. 1996; Michel and Westhof 1990; Westhof and Massire 2004; Przybilski et al. 2005).

What do we currently know about the function of ES? On the small ribosomal subunit (SSU) of protozoan ribosomes, for instance, a network of three ES (ES3S, ES6S, ES7S) span >150 Å in the vicinity of the mRNA exit channel (Gao et al. 2005; Hashem et al. 2013b). This cluster of ES could to be involved in translation initiation through the regulation of recruitment of initiation factors such as eIF3 (Hashem et al. 2013a, b). We currently hypothesize that this platform is required for translation of protozoan mRNAs, which have particular features compared to most know eukaryotic mRNAs (a hypermodified cap, presence of a 39-nt leader sequence

that is trans-spliced to the 5′ UTR of all kinetoplastid mature mRNAs) (Gao et al. 2005; Hashem et al. 2013b). ES may also participate in ribosome biogenesis, as illustrated by the finding that 12 out of the 16 ESs on the large subunit (LSU) of the yeast ribosome are required for optimal growth and correct 25S rRNA processing (Ramesh and Woolford 2016). The roles associated with certain ES in a particular organism may be conserved across various species. For instance, ES27L on the LSU was shown to be consistently alternating between two positions (near L1 and the peptide tunnel exit) from the structures of the human, wheat, fly and yeast ribosomes (Anger et al. 2013; Armache et al. 2010; Beckmann et al. 2001). This observation suggests a similar role, especially since replacing ES27L with that of another species restores function (Sweeney et al. 1994).

In this chapter, we focus on ribosomal ES from the Trypanosoma genus (ES in other eukaryotes have been discussed elsewhere Anger et al. 2013; Li et al. 2017). Tritryps kinetoplastid parasites, i.e. *Trypanosoma cruzi, Trypanosoma brucei* spp. and *Leishmania* spp. are human pathogens causing the Chagas disease, African trypanosomiasis and leishmaniasis, respectively. Chagas disease alone affects eight million people in the Americas (PAHO 2017), and due to migration flows it has now emerged in non-endemic countries such as Spain and the United States [>300,000 infected individuals in the US (Bern and Montgomery 2009)] (Gascon et al. 2010). Current treatments rely on toxic drugs that are often effective only during specific stages of the disease, further highlighting the urge to develop new strategies for targeting these parasites (Field et al. 2017). Being highly distinctive from those of mammalian ribosomes (Gao et al. 2005; Hashem et al. 2013b), theses ES in kinetoplastids could potentially be promising targets for therapeutic applications. However, efficiently targeting kinetoplastid-specific ES requires a deep structural understanding of their organization and their function, which is the motivation behind the comparative analysis we present here.

Expansion Segments in Kinetoplastid Ribosomes Are Large but Structurally Organized

Biochemical and structural studies on ribosomes from kinetoplastids pinpointed unique features for two *Trypanosoma* species. The first structure of the ribosome from *T. cruzi* was solved at 12 Å resolution (Gao et al. 2005) and that of *T. brucei brucei* (referred to as *T. brucei* in the following) at 5.6 Å resolution (Hashem et al. 2013b). We can now compare these structures to higher-resolution structures of the 80S ribosome from *L. donovani* (Zhang et al. 2016; Shalev-Benami et al. 2016, 2017) and *T. cruzi* (Liu et al. 2016; Querido et al. 2017), which we revised and/or expanded from the published models (Fig. 13.1a, b). We also include in this analysis a previously unpublished structure of the *L. major* 80S ribosome (Fig. 13.1c) at ~5 Å resolution (the structure can be shared upon simple request. Images were collected on an FEI Polara electron microscope equipped with a Gatan Summit

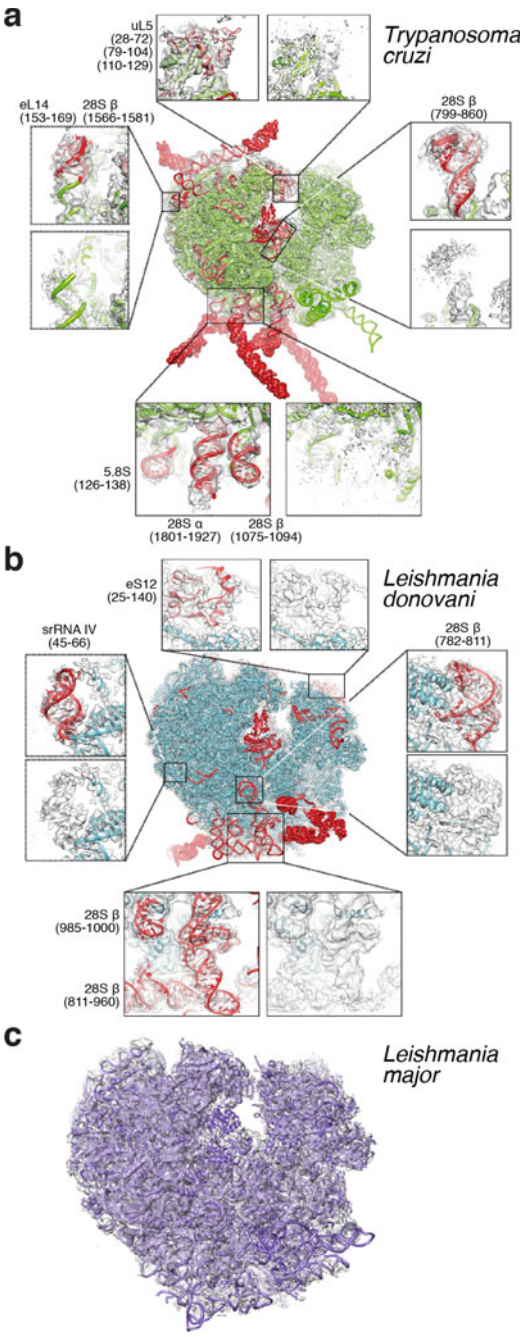


Fig. 13.1 Strategy for modeling flexible expansion segments in ribosomes. **a** Modeling the complete *T. cruzi* 80S at 4 Å resolution. We modeled expansion segments in the SSU of our previously published model [PDB ID: 5OPT Brito Querido et al. 2017; Pitula et al. 2002; Ayub et al. 2012)]. We also included coordinates from a 2.5 Å-structure of the 60S particle from *T. cruzi* [PDB ID: 5T5H (Liu et al. 2016)]. RNA and protein elements were added (a total of 1816 nucleotides and 810 aminoacids) are shown in red. Residues built within density limits (~8 to 15 Å) are shown in ribbon mode, while residues beyond are shown as surfaces. Insets illustrate representative regions of the original unfiltered maps. Dashed boxes point to regions located on the side of the ribosome hidden from view. **b** The quality of the *L. donovani* electron density map at 2.9 Å resolution [PDB ID: 5T2A (Zhang et al. 2016)] enabled us to add 873 nucleotides and 802 aminoacids to the original model (in red). **c** The revised model of the *L. donovani* ribosome was used as the basis for building the structure of the ribosome from *L. major* at 5 Å resolution. The *L. major* map and coordinates will be deposited in the PDB

K2 Camera. After image processing, nearly 150,000 particle images were used in the final reconstruction). Sequence conservation between the *L. donovani* and *L. major* ribosomes is >98% (rRNA) and >97% (r-protein; except for eS24 ~ 93%). Our strategy was to model the ES for these four species of kinetoplastids according to rRNA structural alignments that we generated using non-redundant sequences in SSU-Align (Nawrocki 2009), which we manually adjusted in Ugene (Okonechnikov et al. 2012) and validated using R-scape (Rivas et al. 2017). Because of their high level of flexibility, in many cases it was not possible to directly model some ES in their cryo-EM map densities (Fig. 13.1a, b), as had been done previously for other eukaryotic ribosomes (Anger et al. 2013).

A side-by-side comparison of the ribosomes from *T. brucei*, *L. donovani*, *L. major*, and *T. cruzi* reveals that many ES are several-fold larger than their counterparts in mammals (Fig. 13.2). Although the 28S rRNA in the LSU of *T. brucei* and *T. cruzi* is halfway in size between yeast and human, their 18S in the SSU is >20% longer than their human counterpart (Table 13.1). Most of this extra rRNA sequence is found within ES6S and ES7S that are inserted in helices 21 and 26. ES6S and ES7S are respectively 230 and 17 nucleotides long in *H. sapiens*, but 429 and 136 and 512 and 164 nucleotides long in *L. major* and *T. cruzi*. ES6S, ES7S. It should also be noted that ES9S and ES31L adopt kinetoplastid-specific tertiary structures and they cluster around the mRNA channel exit. In particular, ES6S, ES7S and ES9S are located at a very strategic region where initiation factor 3 is known to bind in mammals, forming bridges between the LSU and SSU that are not found in other eukaryotic ribosomes (Hashem et al. 2013b).

From aligning more than 80 sequences for the SSU and 25 for the LSU, we found that all ES from kinetoplastids have a conserved structure, with large extensions for some helical regions (Figs. 13.3 and 13.4). This confirms and expands on prior findings obtained with fewer than ten kinetoplastid sequences for all ES (Hashem et al. 2013b), and with 3000 eukaryotic sequences for ES6S only (Wuyts et al. 2000). The proportion of conserved versus variable regions varies among ES, with ES3S (~200 nt long) being the most conserved. Large size variations are observed even among kinetoplastids. For example, ES6S, ES7S, ES7L, ES19L and ES27L have a comparable size among the genera *Angomonas*, *Herpetomonas* and *Phytomonas*. The same ES are significantly larger in *Leishmania* and *Trypanosoma* species, the largest being found in *T. cruzi*. Although ES3S, ES9S, ES10S and ES31L are all larger in *Leishmania* spp. and in *T. brucei* than in *T. cruzi*, the size differences remain relatively modest (Table 13.1).

Expansion Segments Form Kinetoplastid-Specific Interaction Networks

The overall high degree of ES conservation can be explained by interaction networks at the ribosome surface. For example, in all cryo-EM structures of kinetoplastids,

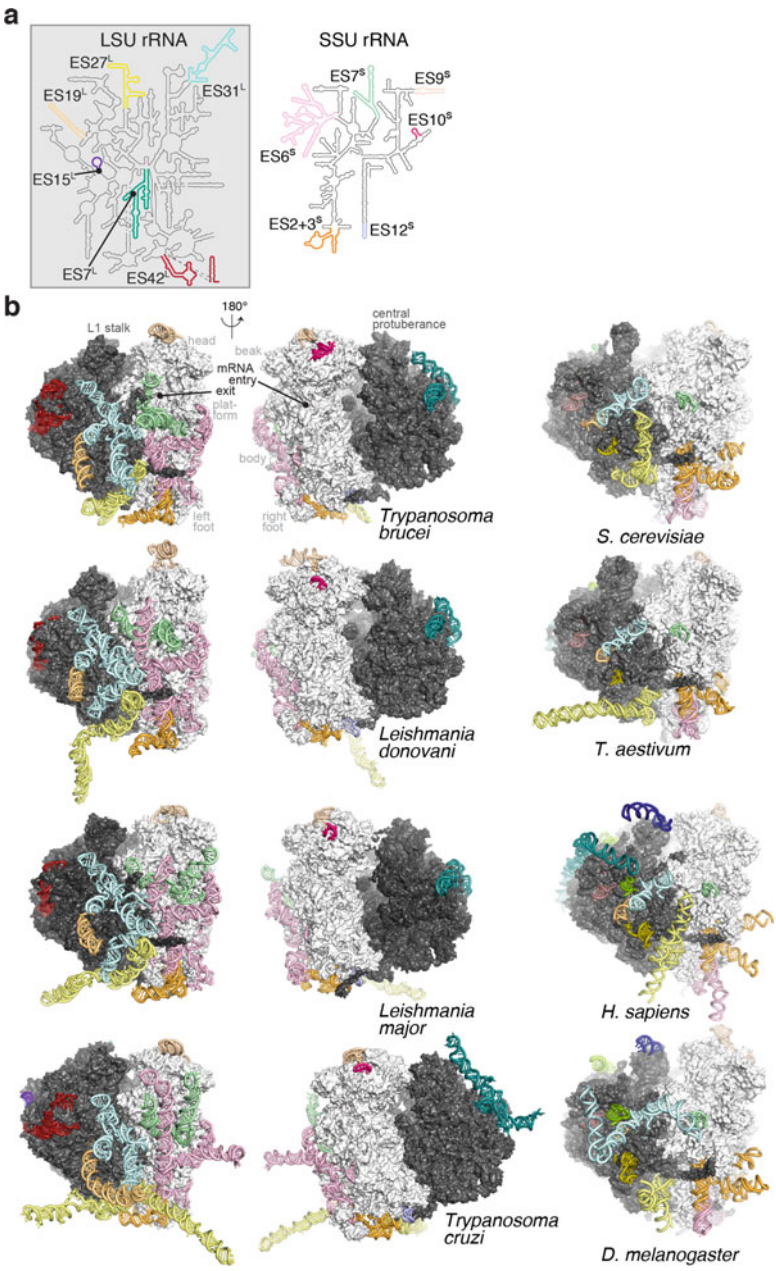


Fig. 13.2 The complex outer layer of expansion segments at the surface of ribosomes from kinetoplastids. **a** Secondary structure schematic for the LSU and SSU rRNA of kinetoplastids. Expansion segments are highlighted in colors and labeled according to conventional numbering (Gerbi 1996). **b** Overall structure of the 80S ribosomes from four kinetoplastids, viewed from the mRNA L1-stalk side (left) and SSU beak sides (middle). Ribosomes are shown in surface view, except for ES rendered as ribbons and colored as in panel (**a**). (Right) Ribosome structures from other organisms are shown from the L1-stalk side as a reference (references in the main text)

Table 13.1 Size comparison of SSU and LSU expansion segments

	L. donovani	*L. major*	*T. cruzi*	*T. brucei*	*S. cerevisiae*	*D. melanogaster*	*H. sapiens*
ES2S	22	22	22	22	18	16	21
ES3S	147	147	148	156	113	121	159
ES6S	429	429	512	406	222	300	230
ES7S	138	136	164	159	17	17	17
ES9S	76	76	76	89	45	151	52
ES10S	17	20	20	29	5	5	7
ES12S	25	25	26	26	38	41	43
Total for SSU	854	855	968	887	458	651	529
ES5L	42	42	47	47	45	50	50
ES7L	203	203	291	219	200	331	866
ES15L	32	30	39	32	15	41	189
ES19L	58	58	124	63	29	25	39
ES27L	213	213	323	234	159	222	714
ES31L	237	236	220	235	70	208	85
Total for LSU	785	782	1044	830	518	877	1893
Total number of ES ntds	1639	1637	2012	1717	976	1528	2422

The number of nucleotides is indicated for every conserved ES in seven eukaryotic species

ES3S is tightly packed as it forms the left foot of the SSU (Fig. 13.3a). ES3S interacts with ES2S, domains B and D of ES6S (Fig. 13.4a), ES12S (forming the right foot of the SSU) and r-proteins eS4, eS6, eS8 and KSRP (Fig. 13.3a, b). The structure of the pseudoknot formed between ES3S and ES6S is conserved across parasitic and eukaryotic ribosomes of known structures, as reported earlier (Armache et al. 2010; Alkemar and Nygard 2006; Hashem et al. 2013b; Zhang et al. 2016). The kinetoplastid-specific r-protein (KSRP) that we identified in *T. cruzi* (Brito Querido et al. 2017) is similarly bound to ES3S and ES6S in *Trypanosoma* and *Leishmania,* through its two conserved RNA recognition motifs (RRM) and a C-terminal helix that also interacts with eS6 (Fig. 13.3c). The only variable region within ES3S is the Ab stem, which nonetheless conserves its spatial arrangement as it bends around eS8 in the four kinetoplastids models (Fig. 13.3d). It is important to emphasize that ES3S simultaneously displays structural characteristics conserved in all eukaryotes (pseudoknot with ES6S) and unique features only found in kinetoplastids (KSRP binding).

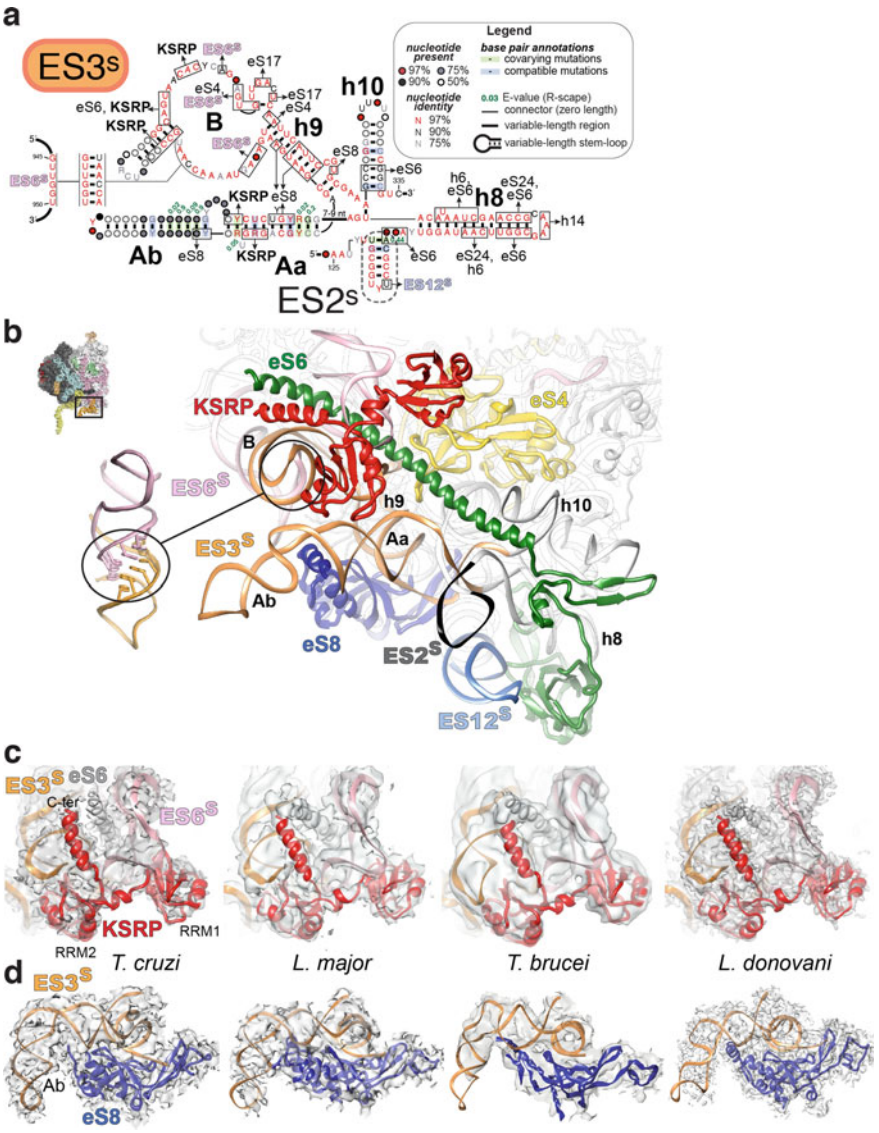


Fig. 13.3 Conservation of interaction networks between kinetoplastids. **a** Interaction network and consensus secondary structure of ES3S. Note that ES2S forms a hairpin at the 5′ of h8. Statistically significant covarying base pairs are shown in green [E-value <0.05 (Rivas et al. 2017)]. This and subsequent covariation models were generated using R2R (Weinberg and Breaker 2011). **b** Close-up of the left foot of the ribosome from *L. donovani*, showing the arrangement of several ES and r-proteins. **c** Conserved binding of KSRP, which interacts with ES3S, ES6S and eS6. **d** Conserved spatial arrangement of stem Ab of ES3S and r-protein eS8. Sequence alignments for all ES are available directly from the authors upon request

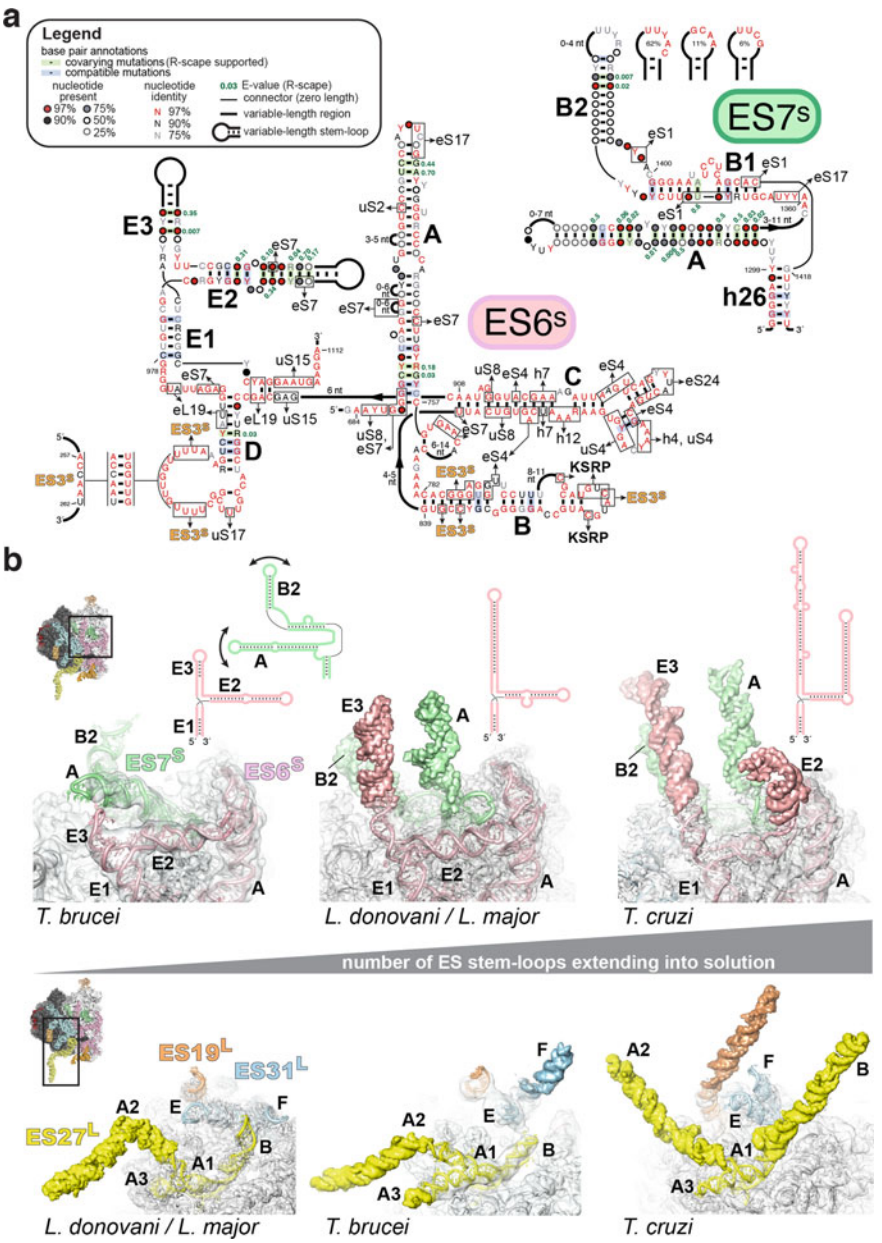


Fig. 13.4 Sequence and structure variability among ES is significant but localized. **a** Interaction networks and consensus secondary structures of ES6S and ES7S (color-coding as in Fig. 13.3). **b** Conformational variability as a function of the length of certain ES across kinetoplastids. Top: The size of ES7S barely changes across kinetoplastids, in contrast to that of ES6S. The increasing size of ES6S leads to displacement of ES7S helices. The regions covered by the electron density maps (grey) are shown as ribbons, while the sections modeled beyond the limit of the current experimental maps are shown as surfaces. Bottom: Conformational changes in particular of ES31L occur as the length of the surrounding ES19L and ES27L increases

Even for larger and more variable ES such as ES6S and ES7S, their core shows a high degree of conservation (Fig. 13.4a), supporting a conserved topology. ES6S is the largest ES on the SSU (400–500 nt) and its highly conserved domains A–D interact tightly with r-proteins and KSRP at the SSU solvent-exposed surface. ES6S interacts also with ES7S, which is of a similar size in *T. cruzi* and *T. brucei*, and ~25% larger in *T. cruzi* than in *L. donovani*. In contrast to stem Ab in ES3S, whose length changes only slightly across parasites, stems E1–E3 in ES6S roughly double in size from *T. brucei* through *Leishmania* to *T. cruzi* (Fig. 13.4a, b). As a result, the number of stem-loops protruding into the solvent is different among parasites (Fig. 13.4b). Both the nature and the presence of these stems dictate the positioning of the conserved stems of ES7S, which are possibly more conformationally restricted in *Leishmania* and *T. cruzi* than in *T. brucei*. Note that although the lack of density prevented pinpointing the exact location of the expansions seen in E2 and E3, the interaction between E3 and ES7S seen in *T. brucei* cannot occur for the other three model parasites, where E3 is much longer (Fig. 13.4b). These structural differences probably reflect the dynamic properties of this area, which is a 'hotspot' for translation initiation regulation (Hashem et al. 2013a).

ES 9S, 10S and 12S, on the other hand, are among the smallest SSU ES. ES12S, for example, is an extension of h44 by 9–10 base pairs in all kinetolastids, while ES10S exists as a 5–9 base pair stem only in *T. brucei, T. congolense, T. simiae, and T. vivax* (Fig. 13.5). The shorter ES seem to follow a similar trend as the longer ES, thus presenting conserved cores and variable stems.

On the LSU, ES are concentrated on two regions, the L1-stalk region, and the solvent-exposed face (Fig. 13.2b). The ES at the solvent-exposed face appear less interconnected compared to the SSU ES. ES7L is isolated from the other sizable LSU ES. It presents a two-pronged fork shape (Figs. 13.6a and 13.7c). Because of its numerous interactions with r-proteins (Fig. 13.6a), its densities are relatively well resolved and present a good local resolution. It forms a pseudoknot with the small ES15L (Fig. 13.6b, d). Another more isolated ES is 5L (also called ES42L) (Fig. 13.6e). Although situated close to ES7L, no direct interactions were observed with the latter. ES42L presents a high level of sequence conservation and interacts with r-proteins eL8, eL15 and uL29 (Fig. 13.6e).

On the other hand, LSU ES near the L1-stalk region are interconnected, where ES19L, ES27L and ES31L are clustered (Figs. 13.2b and 13.6). The length of ES31L varies only within a range of ~20 nucleotides among kinetoplastids, while that of ES19L and ES27L increases by ~50% between *L. donovani* and *T. cruzi* (Table 13.1). As a result, stem F of ES31L adopts different orientations in the various structures, where it is constricted to various degrees by nearby ES (Fig. 13.4b). ES31L is highly structured, as it comprises four nested three-way junctions but a high degree of sequence variability (Fig. 13.6h). Although ES31L possesses a complex predicted tertiary structure, its cryo-EM corresponding densities are poorly resolved in all four kinetoplastid model ribosomes, probably because of its high flexibility. Although ES31L interacts with r-proteins eL27 and eL34 (Fig. 13.6h), the bulk of its rRNA is not stabilized by any other r-proteins, which explains its high level of flexibility. The cryo-EM reconstructions from several kinetoplatids strongly suggest its interaction

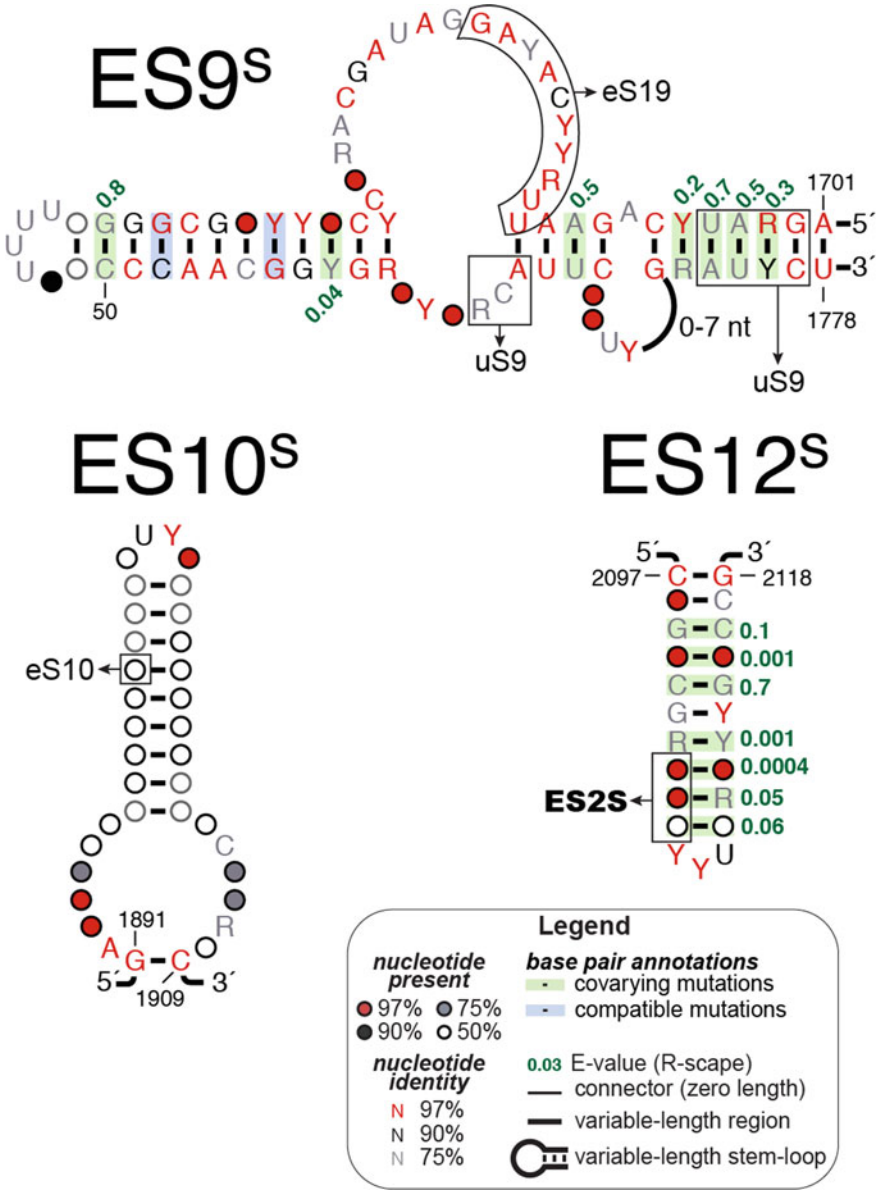


Fig. 13.5 Covariation analysis of SSU expansion segments. Consensus sequences and secondary structures for ES9S, ES10S and ES12S, with interacting partners as on Figs. 13.3 and 13.4

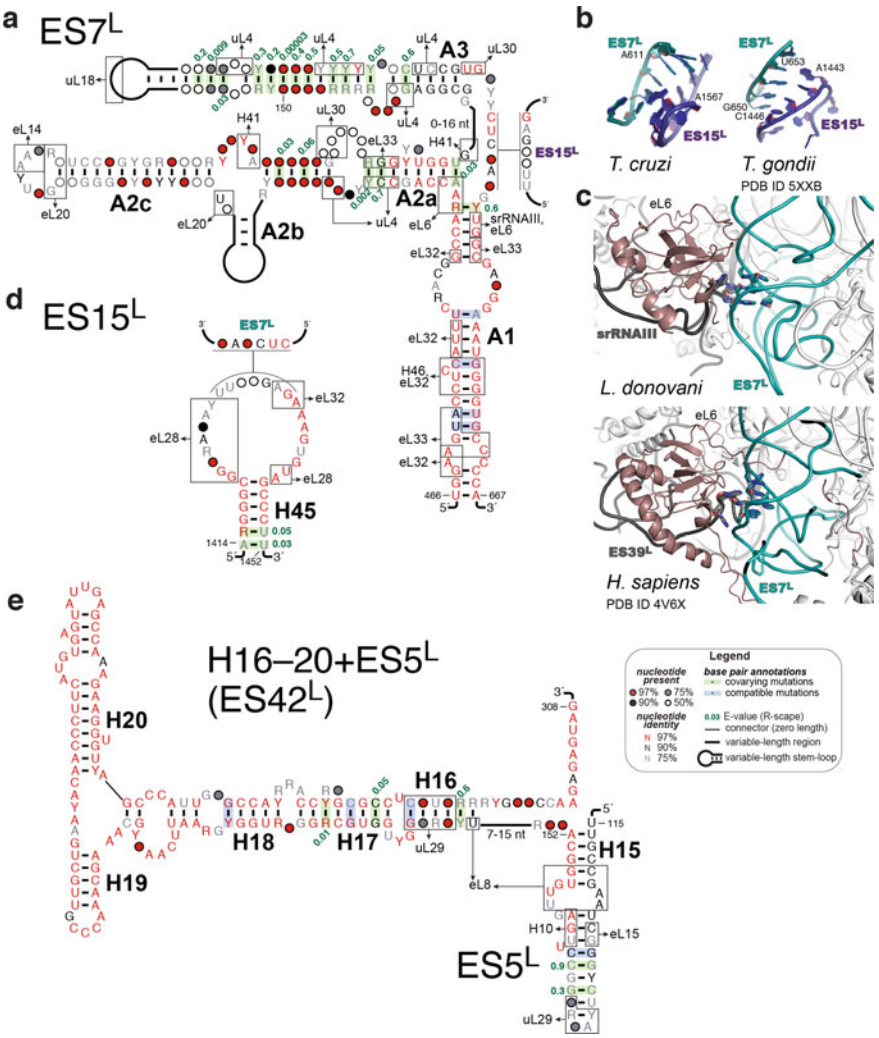


Fig. 13.6 Covariation analysis and interaction networks of LSU expansion segments. **a** Consensus sequence and secondary structure for ES7L. **b** Close-up of the pseudoknot formed by ES7L and ES15L in *T. cruzi* and *T. gondii* [PDB ID: 5XXB (Li et al. 2017)]. **c** Close-up of the region at the top of stem A1 of ES7L in *L. donovani* and *H. sapiens*. **d–f** Consensus sequences and secondary structures for ES15L, ES5L (also known as ES42L) and ES19L (interacting partners are not shown for H16–20). **g** Close-up of the region comprising ES19L and ES31L. **h–i** Same as (**d–f**) for ES31L and ES27L

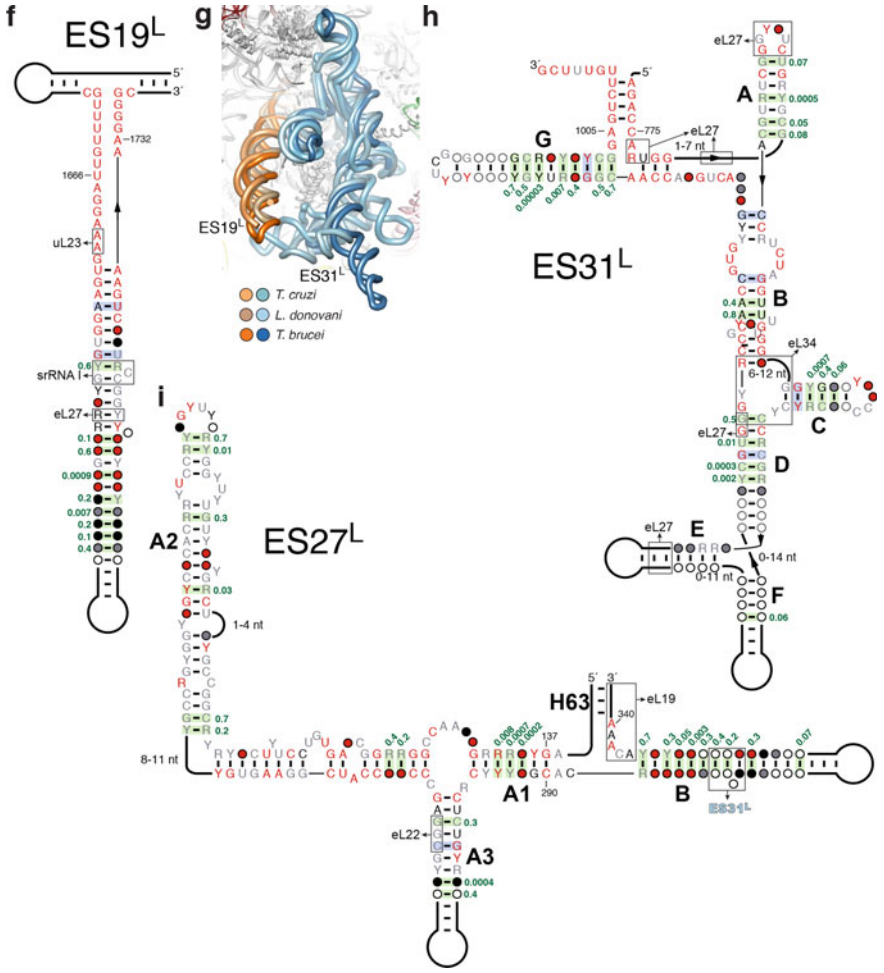


Fig. 13.6 (continued)

with ES19L and ES27L because of its proximity to the latter ES (Figs. 13.2b and 13.6g). ES19L appears as an elongated stem of a relatively well resolved mass of density that shows its interactions with r-proteins eS27 and uL23, but also with srRNA1 (Fig. 13.6f). It is interesting to note that ES31L and ES19L interaction occurs indirectly through r-protein eL27 (Fig. 13.6f, h). Last but not least, ES27L contains two three-way junctions conferring it an elongated shape with two grafted branches protruding toward the solvent (Fig. 13.6i). It interacts directly with ES31L and r-proteins eL19 and eL22 (Fig. 13.6i). Thus, all three, ES 19L, 27L and 31L form an interconnected bundle of rRNA near the L1-stalk on the LSU (Fig. 13.2).

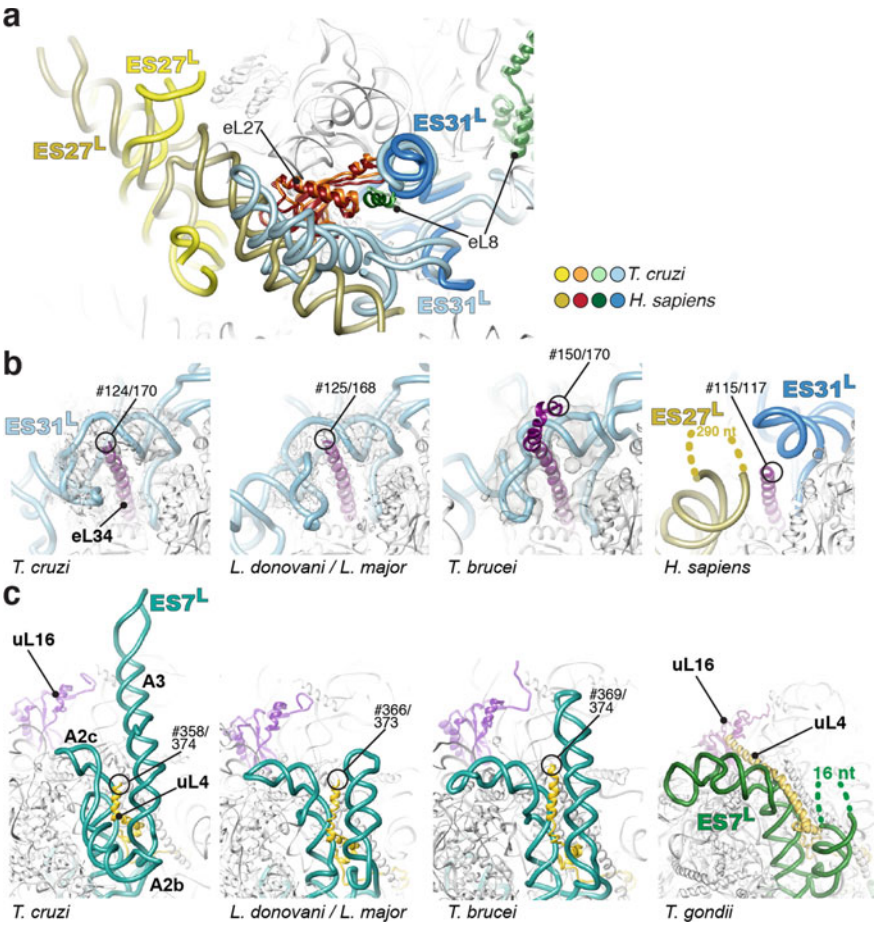


Fig. 13.7 Compensatory interaction networks on the LSU. **a** Superimposition of the ribosomes from *T. cruzi* (light shades) and *H. sapiens* [PDB ID: 4V6X; dark shades (Anger et al. 2013)]. The view is from the top of that shown in panel 22.4b, after removing regions modeled outside of density. **b** Close-up of the location where eL34 binds ES31L in kinetoplastids, and of the corresponding region in the human ribosome (PDB ID: 4V6X). For kinetoplastids, the number of the last residue modeled in density as well as the total length of the r-protein are indicated. **c** Close-up of ES7L bound to uL4 in kinetolastids, and of the corresponding region in the *T. gondii* ribosome [PDB ID: 5XXB (Li et al. 2017)], where an elongated uL4 interacts with uL16

Finally, it is worth mentioning that expansion segments may have a conserved structure even in the absence of sequence conservation, as exemplified by ES7S (Figs. 13.2 and 13.4b), and subdomain A3 of ES7L (Figs. 13.6a and 13.7c).

Compensatory Interaction Networks

Compensatory interaction networks are widespread on the LSU, as previously observed upon comparing ribosomes from yeast, wheat, fly, human and several parasites (Anger et al. 2013; Armache et al. 2010; Li et al. 2017; Hashem et al. 2013b). For instance, the larger ES31L in kinetoplastids (70 nt in yeast, 85 nt in human, but 220–237 nt in kinetoplastids) fills the space occupied by an extended helix from the 714 nt-long ES27L in humans, without altering the binding mode of eL27 and eL8 (Fig. 13.7a). Another example is the conserved stem A1 of ES7L that packs against eL6, eL32 and eL33 but not eL28, in contrast to ES7L of other eukaryotic ribosomes (Anger et al. 2013; Ben-Shem et al. 2011; Hashem et al. 2013b; Li et al. 2017) (Fig. 13.6a, d). Moreover, although the pseudoknot formed between ES7L and ES15L is similar to that in *T. gondii* (Fig. 13.6b, d), the long-range interaction involving the shallow groove of ES7L and a loop from ES39L in human is replaced by interactions with srRNA3 (Fig. 13.6c).

Size increase of ES is often accompanied by extensions of r-proteins (Hashem et al. 2013b; Li et al. 2017). For example, the eL34 r-protein is extended in kinetoplastids by ~50 amino acids compared with human eL34, which can be explained by a larger ES31L (Fig. 13.7b). It therefore came as a surprise that the length of uL4 remains the same across kinetoplastids, even though stem A3 of ES7L nearby dramatically varies (Fig. 13.7c). Nevertheless, exceptions can be observed, such as in *P. falciparum* and *T. gondii* where ES7L is of a similar size compared to that of *T. brucei* (~290 nt). However, uL4 is elongated by ~40 amino acids in *P. falciparum* and *T. gondii*, so that it interacts with uL16 (Fig. 13.7c), a late binding protein during ribosome assembly in yeast (Li et al. 2017; Ma et al. 2017), with no obvious correlation with the size of ES7L (Fig. 13.7c).

Concluding Remarks

In line with previous work that presented the structures, and underscored the importance of, expansion segments in other species (Armache et al. 2010; Beckmann et al. 2001; Ben-Shem et al. 2011), this chapter illustrates that ES are not random extensions with poor sequence conservation at the surface of ribosomes. Rather, ES are organized so that they tightly interact with core rRNA and r-proteins, recruit kinetoplastid-specific proteins, and act as dynamic gatekeepers. Some of these interaction networks are found in some but not all parasites, suggesting that the precise role of ES may vary even among species from the same phylum. Through highlighting the importance of ES networks in the ribosomes of kinetoplastid parasites,

this work expands the chacterization of kinetoplastid-specific features, in the hope of stimulating the development of safer and more specific anti-parasitic therapeutic.

Authors Contributions

QV and AB did the molecular modeling and cryo-EM maps (re)analysis. AJ and YH acquired *L. major* ribosome particles and processed the cryo-EM data. QV, YH and JF wrote the manuscript with input from all authors. QV and AB did the Fig.s. YH and JF supervised the research.

Acknowledgements We thank Robert A. Grassucci for assistance in data acquisition of *L. major*. This work was supported by the ANR grant ANR-14-ACHN-0024 @RAction program "ANR CryoEM80S" and the ERC-2017-STG #759120 "TransTryp" (to Y.H.). J.F. acknowledges support by HHMI and NIH R01 GM29169.

References

Alkemar G, Nygard O (2006) Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. Biochemistry 45(26):8067–8078. https://doi.org/10.1021/bi052149z

Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R (2013) Structures of the human and Drosophila 80S ribosome. Nature 497(7447):80–85. https://doi.org/10.1038/nature12104

Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Soding J, Westhof E, Wilson DN, Beckmann R (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. Proc Natl Acad Sci USA 107(46):19748–19753. https://doi.org/10.1073/pnas.1009999107

Ayub AJ, Lapadula WJ, Hoebeke J, Smulski CR (2012) Ribosomes from Trypanosomatids: unique structural and functional properties. In: Biyani M (ed) Cell-free protein synthesis. Intech, pp 3–28

Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J, Blobel G (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell 107(3):361–372

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. Science, 334. https://doi.org/10.1126/science.1212642

Bern C, Montgomery SP (2009) An estimate of the burden of Chagas disease in the United States. Clin Infect Dis 49(5):e52–54. https://doi.org/10.1086/605091

Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD (2018) Translation: the universal structural core of life. Mol Biol Evol 35(8):2065–2076. https://doi.org/10.1093/molbev/msy101

Brito Querido J, Mancera-Martínez E, Vicens Q, Bochler A, Chicher J, Simonetti A, Hashem Y (2017) The cryo-EM structure of a novel 40S kinetoplastid-specific ribosomal protein. Structure 25(12):1785–1794.e3. https://doi.org/10.1016/j.str.2017.09.014

Clark CG, Tague BW, Ware VC, Gerbi SA (1984) Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res 12(15):6197–6220. https://doi.org/10.1093/nar/12.15.6197

Field MC, Horn D, Fairlamb AH, Ferguson MAJ, Gray DW, Read KD, De Rycker M, Torrie LS, Wyatt PG, Wyllie S, Gilbert IH (2017) Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. Nat Rev Microbiol 15(7):447. https://doi.org/10.1038/nrmicro.2017.69

Gao H, Ayub MJ, Levin MJ, Frank J (2005) The structure of the 80S ribosome from Trypanosoma cruzi reveals unique rRNA components. Proc Natl Acad Sci USA 102(29):10206–10211. https://doi.org/10.1073/pnas.0500926102

Gascon J, Bern C, Pinazo MJ (2010) Chagas disease in Spain, The United States and other non-endemic countries. Acta Trop 115(1–2):22–27. https://doi.org/10.1016/j.actatropica.2009.07.019

Gerbi SA (1996) Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. In: Zimmermann RA, Dahlberg AE (eds) Ribosomal RNA: Structure. Telford - CRC Press, Evolution, processing and function in protein synthesis, pp 71–87

Hashem Y, des Georges A, Dhote V, Langlois R, Liao HY, Grassucci RA, Hellen CU, Pestova TV, Frank J (2013a) Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153(5):1108–1119. https://doi.org/10.1016/j.cell.2013.04.036

Hashem Y, des Georges A, Fu J, Buss SN, Jossinet F, Jobe A, Zhang Q, Liao HY, Grassucci RA, Bajaj C, Westhof E, Madison-Antenucci S, Frank J (2013b) High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome. Nature 494(7437):385–389. https://doi.org/10.1038/nature11872

Hassouna N, Michot B, Bachellerie JP (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res 12(8):3563–3583. https://doi.org/10.1093/nar/12.8.3563

Kachouri R, Stribinskis V, Zhu Y, Ramos KS, Westhof E, Li Y (2005) A surprisingly large RNase P RNA in Candida glabrata. RNA 11(7):1064–1072. https://doi.org/10.1261/rna.2130705

Lehnert V, Jaeger L, Michel F, Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. Chem Biol 3(12):993–1009

Li Z, Guo Q, Zheng L, Ji Y, Xie YT, Lai DH, Lun ZR, Suo X, Gao N (2017) Cryo-EM structures of the 80S ribosomes from human parasites *Trichomonas vaginalis* and *Toxoplasma gondii*. Cell Res 27(10):1275–1288. https://doi.org/10.1038/cr.2017.104

Liu Z, Gutierrez-Vargas C, Wei J, Grassucci RA, Ramesh M, Espina N, Sun M, Tutuncuoglu B, Madison-Antenucci S, Woolford JL Jr, Tong L, Frank J (2016) Structure and assembly model for the *Trypanosoma cruzi* 60S ribosomal subunit. Proc Natl Acad Sci USA 113(43):12174–12179. https://doi.org/10.1073/pnas.1614594113

Ma C, Wu S, Li N, Chen Y, Yan K, Li Z, Zheng L, Lei J, Woolford JL Jr, Gao N (2017) Structural snapshot of cytoplasmic pre-60S ribosomal particles bound by Nmd3, Lsg1, Tif6 and Reh1. Nat Struct Mol Biol 24:214–220

Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M (2012) One core, two shells: bacterial and eukaryotic ribosomes. Nat Struct Mol Biol 19(6):560–567. https://doi.org/10.1038/nsmb.2313

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216(3):585–610. https://doi.org/10.1016/0022-2836(90)90386-Z

Nawrocki EP (2009) Structural RNA homology search and alignment using covariance models. Washington University School of Medicine. https://doi.org/10.7936/K78050MP

Okonechnikov K, Golosova O, Fursov M, Team U (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28(8):1166–1167. https://doi.org/10.1093/bioinformatics/bts091

PAHO (2017) Chagas disease. Pan American Health Organization. https://www.paho.org/hq/index.php?option=com_topics&view=article&id=10&Itemid=40743. Accessed 20 Oct 2017

Pitula J, Ruyechan WT, Williams N (2002) Two novel RNA binding proteins from *Trypanosoma brucei* are associated with 5S rRNA. Biochem Biophys Res Commun 290(1):569–576. https://doi.org/10.1006/bbrc.2001.6226

Przybilski R, Graf S, Lescoute A, Nellen W, Westhof E, Steger G, Hammann C (2005) Functional hammerhead ribozymes naturally encoded in the genome of *Arabidopsis thaliana*. Plant Cell 17(7):1877–1885. https://doi.org/10.1105/tpc.105.032730

Ramesh M, Woolford JL Jr (2016) Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. RNA 22(8):1153–1162. https://doi.org/10.1261/rna.056705.116

Rivas E, Clements J, Eddy SR (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods 14(1):45–48. https://doi.org/10.1038/nmeth.4066

Shalev-Benami M, Zhang Y, Matzov D, Halfon Y, Zackay A, Rozenberg H, Zimmerman E, Bashan A, Jaffe CL, Yonath A, Skiniotis G (2016) 2.8-A Cryo-EM structure of the large ribosomal subunit from the eukaryotic parasite leishmania. Cell Rep 16 (2):288–294. https://doi.org/10.1016/j.celrep.2016.06.014

Shalev-Benami M, Zhang Y, Rozenberg H, Nobe Y, Taoka M, Matzov D, Zimmerman E, Bashan A, Isobe T, Jaffe CL, Yonath A, Skiniotis G (2017) Atomic resolution snapshot of Leishmania ribosome inhibition by the aminoglycoside paromomycin. Nat Commun 8(1):1589. https://doi.org/10.1038/s41467-017-01664-4

Sweeney R, Chen L, Yao MC (1994) An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. Mol Cell Biol 14(6):4203–4215

Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, Gerbi SA (1983) Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis. Nucleic Acids Res 11(22):7795–7817. https://doi.org/10.1093/nar/11.22.7795

Weinberg Z, Breaker RR (2011) R2R–software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinf 12:3. https://doi.org/10.1186/1471-2105-12-3

Westhof E, Massire C (2004) Structural biology. Evolution of RNA architecture. Science 306(5693):62–63. https://doi.org/10.1126/science.1104482

Wuyts J, De Rijk P, Van de Peer Y, Pison G, Rousseeuw P, De Wachter R (2000) Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. Nucleic Acids Res 28(23):4698–4708

Zhang X, Lai M, Chang W, Yu I, Ding K, Mrazek J, Ng HL, Yang OO, Maslov DA, Zhou ZH (2016) Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nat Commun 7:13223. https://doi.org/10.1038/ncomms13223

Chapter 14
Hepatitis B Core Protein Capsids

Bettina Böttcher

Abstract Non-enveloped Nackednaviridae and enveloped hepadnaviridae both have capsids that are formed by related small proteins which evolved more than 430 Mya. In Hepatitis B virus, which belongs to the enveloped hepadnaviridae, this small protein is termed Hepatitis B core protein (Hbc). Its function, as building block of a major human pathogen, triggered extensive research that elucidated the importance of almost every single amino acid for the structural integrity of the capsids and the orderly progression of the viral life cycle. In particular, encapsidation of the genome, envelopment of the capsid, uncoating of the genome and targeting of the different compartments during viral maturation have been a vivid focus of research. HBc has also been developed as a biotechnological tool for the design of nano-containers with tailored properties. These nano-containers can display foreign epitopes on their surfaces and induce a strong immune response, which is attractive for the development of vaccines against other pathogens. This chapter will discuss some of the unique properties of HBc and their significance for the formation of a functional macromolecular capsid.

Keywords Hepatitis B virus · Hepatitis B core protein · Icosahedral capsid · Spike · Nano-container · C-terminal domain · Assembly domain · Arginine-rich domain

Hepatitis B Core Protein a Component of the Hepatitis B Virus

Hepatitis B Virus (HBV) belongs to the family of Hepadnaviridae and is a major human pathogen. Approximately one third of the world population has been in contact with the virus (Alter 2003) and around 250 million have become chronic carriers with an increased risk for developing liver cirrhosis and primary liver cancer. The virions

B. Böttcher (✉)
Rudolf Virchow Center and Biocenter, Julius Maximilians Universität Würzburg, Josef-Schneider Straße 2, 97080 Würzburg, Germany
e-mail: Bettina.bottcher@uni-wuerzburg.de

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_14

consist of two shells. The outer shell is a membrane containing envelope, which is densely packed with surface proteins (HBs). The inner shell is a capsid that is formed entirely by a protein called hepatitis B core protein (HBc) or Hepatitis B core antigen (HBcAg). These HBc subunits assemble around a complex of viral pregenomic (pg) RNA and viral polymerase (protein P) (for review see Nassal 2008; Ning et al. 2017) and form an icosahedral capsid. Inside the capsid the viral polymerase reversely transcribes the RNA into partly double-stranded (ds) DNA. After reverse transcription is completed, capsids are either targeted towards the nucleus (early after infection) or they are enveloped by a lipid membrane, which is densely packed with HBs (Fig. 14.1). These mature virions (Summers and Mason 1982) are secreted together with a huge excess of subviral particles formed by HBs and virus-like, enveloped particles containing empty, highly phosphorylated HBc capsids (Ning et al. 2011).

The viral envelope is densely packed with three types of surface proteins (HBs) which are termed large (L), medium (M) and small (S). All types have of HBs have the same C-terminal membrane-embedded domain that comprises S-HBs and consists of 4 transmembrane helices. M-HBs and L-HBs have N-terminal additions of increasing length. These additions are termed preS2 for the part that is common in M-HBs and L-HBs and preS1, for the part which is unique to L-HBs. The whole N-terminal extension to S-HBs in L-HBs is commonly referred to as the PreS-domain

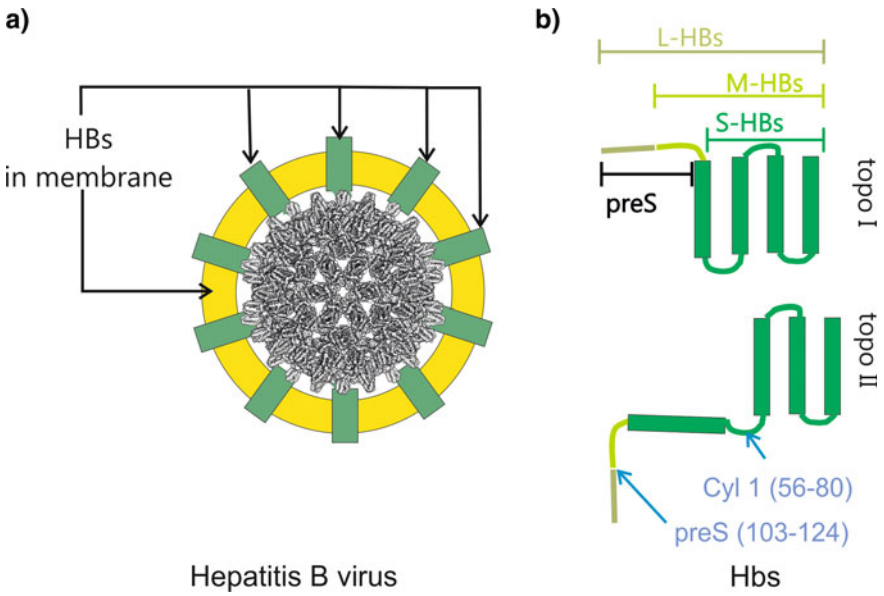


Fig. 14.1 Schematic representation of Hepatitis B Virus (HBV). **a** The viral capsid (grey) is enveloped by a lipid membrane (yellow), which is densely packed with surface proteins (green). **b** The three types of surface proteins (S, L, M-HBs) have increasing extensions to the N-terminus. The largest variant, L-HBs, adopts two topologies (topo I and topo II). Topo I supports binding of L-HBs to the cellular receptor and topo II to the capsid

(Fig. 14.1). For the envelopment of HBc L-HBs and S-HBs are important while M-HBs is dispensable. It is still an ongoing debate where the binding site for these two surface proteins on the capsids is.

In the envelope, L-HBs has a dual function as it binds to the cellular receptor during infection and to the capsid during envelopment. This is achieved by a dual topology (Bruss et al. 1994) in which the N-terminus of L-HBs can either face the luminal or cytosolic site of the membrane (Fig. 14.1, topo I and topo II). Viral secretion by budding requires enveloped particles. Infective particles include capsids with a mature DNA-genome and an envelope that contains L-HBs. All other types of secreted particles are none infective.

Structure of Hepatitis B Core Protein

The capsid forming human HBc is a small protein of only 183–185 amino acid residues. It consists of an N-terminal assembly domain (1–144), which forms the ordered part of the capsid and a C-terminal domain (CTD), which is largely disordered. The CTD is rich in arginines (arginine rich domain, ARD) and contains several phosphorylation sites, a nuclear export signal (NES) and nuclear localization signals (NLS) (Fig. 14.2). Thus the CTD accommodates the main functional sites of HBc that control ordered progression of viral maturation and its localization within the host cell.

The structure of the assembly domain of HBc is known from numerous studies by electron cryomicroscopy (Crowther et al. 1994; Kenney et al. 1995; Bottcher et al. 1997; Yu et al. 2013; Böttcher and Nassal 2018; Conway et al. 1997; Schlicksup et al. 2018) and X-ray crystallography (Alexander et al. 2013; Bourne et al. 2006; Venkatakrishnan et al. 2016; Wynne et al. 1999). The assembly domain is mainly α-helical and contains 5 helices (Bottcher et al. 1997; Wynne et al. 1999), which are numbered from the N- to the C-terminus (I–V). The N-terminal part of the assembly domain (residues 1–48) accommodates two short helices connected by extended strands of protein. This N-terminal part resembles a fulcrum and loops around the central helical hairpin formed by the two longest helices III and IV (residues 48–111). The central helical hairpin is followed by helix V which together with a subsequent proline rich turn and an extended stretch form a hand-like region (Fig. 14.2). Helices IV and V are flexibly linked at Gly 111 (Böttcher et al. 2006; Packianathan et al. 2010), enabling a hinge-like movement between the two domains.

Most structures resolve the assembly domain only up to residue 144 and in some exceptional cases up to residue 152 (Böttcher and Nassal 2018). However, the density downstream of residue 144 is very weak suggesting either flexibility or lower occupancy with some of the C-termini being in different conformations. This region between the ordered well-resolved assembly domain and the unresolved ARD is often referred to as linker-region. The unresolved residues following 152 contain the Arginine rich domain and the phosphorylation sites of which seven can be phosphorylated in one HBc-monomer in vitro (Heger-Stevic et al. 2018a).

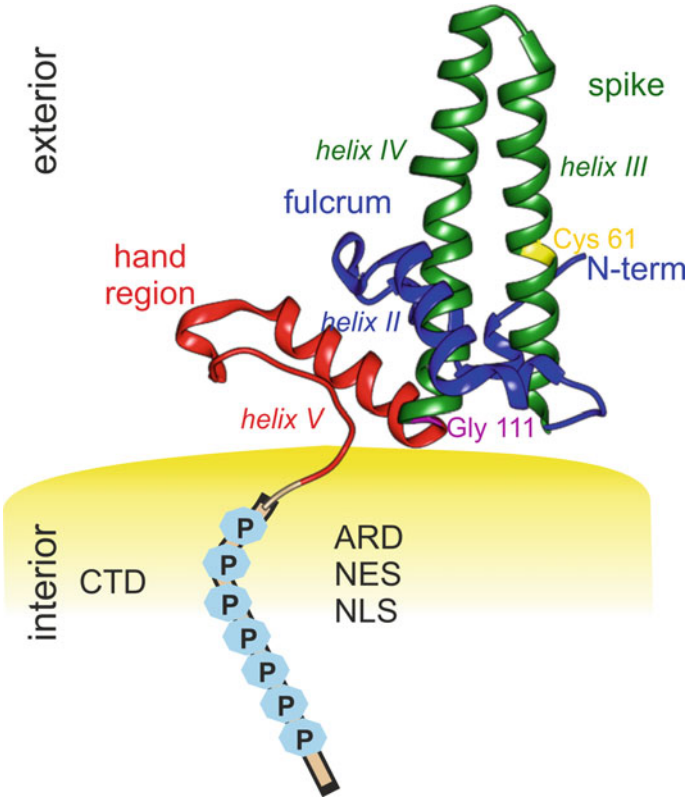


Fig. 14.2 Cartoon of HBc-monomer [chain C of 6HTX (Böttcher and Nassal 2018)]. The assembly domain is shown in colour: N-terminal fulcrum in blue, central spike in green and hand-region in red. The hinge at Gly 111 between the spike and the hand region is indicated in magenta. Cys61 that forms an inner-dimeric SS-bridge in HBc dimers is shown in yellow. The helices are labelled in roman numbers following the numbering of 1QGT (Wynne et al. 1999). Most published structures resolve the assembly domain only up to residue 144. The subsequent CTD is largely disordered and contains seven sites that can be phosphorylated in vitro (Heger-Stevic et al. 2018a), the arginine rich domain (ARD) nuclear localization signals (NLS) and the nuclear export signal (NES). The last resolved residue of HBc in capsids points towards the capsid interior, while the N-terminus is located at the capsid exterior

HBc Forms Dimeric Building Blocks

The Hbc monomer on its own is most likely unstable and has not been observed in solution yet. Folding experiments suggest that two HBc monomers associate as partly unfolded proteins into a dimer and then fold together (Alexander et al. 2013) into a mature dimer. The dimers are hammerhead-shaped (Crowther et al. 1994) with the handle as the inner dimer interface. This interface is formed by helices III and IV of each monomer. The dimer is further stabilized by an inner-dimer SS-bridge

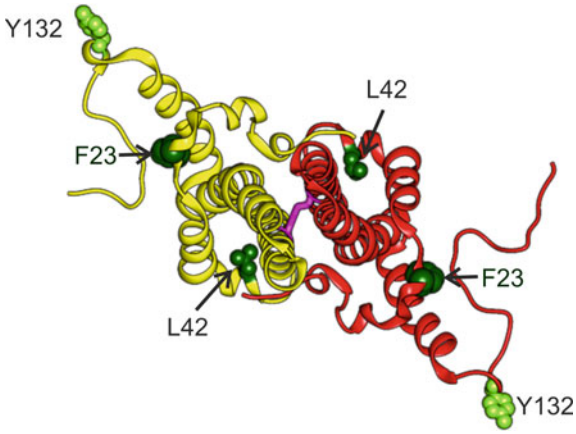


Fig. 14.3 Dimer in cartoon representation (CD dimer of capsids of wt-HBc [6HTX, (Böttcher and Nassal 2018)]. The dimer is shown along the dimer axis from the capsid exterior. The inner dimer SS-bridge in the centre of the spike is shown in magenta. Positions of point mutants that stabilize dimers (L42A, F23A) and trimers of dimers (Y132A) and abolish capsid formation are shown as green balls

(Fig. 14.3, magenta) between Cys 61 (Fig. 14.2, yellow) in the center of helix III. The SS-bridge is dispensable for dimer and capsid formation (Nassal 1992) and slows down capsid assembly in vitro (Selzer et al. 2014).

HBc, C-terminally truncated at residue 138, cannot assemble into capsids and remains in its dimeric form, while longer constructs are able to assemble into capsids under favourable conditions (Zlotnick et al. 1996). The dimeric form is also stabilized over capsids at pH 9.5 and at low ionic strength (Wingfield et al. 1995) and stays dimeric under reducing conditions, even at high protein concentrations (Freund et al. 2008). Certain point mutations abrogate the ability of HBc to assemble into capsids but still support the dimer formation. The probably best-known point mutation that abolishes capsid formation is an exchange of Tyrosine_{132} to Alanine. This position is at the strained turn of the hand region (Fig. 14.3). In human HBc such mutants form stable, planar trimers of dimers whereas in woodchuck HBc the corresponding mutant forms dimers (Zhao et al. 2019). The inability of these mutants to form capsids and their ability to form well diffracting crystals, made them an important construct for routine high-resolution studies of effector binding (Zhou et al. 2017; Qiu et al. 2016; Klumpp et al. 2015) well below 2 Å resolution. Other point mutations in the fulcrum such as F23A and L42A (Fig. 14.3) also abolish capsid formation and lead to dimers that are stable under a wide range of conditions (Alexander et al. 2013).

Hbc Dimers Assemble into Icosahedral Capsids

Dimers are the main building block from which capsids with either T = 3 (180 subunits) or T = 4 (240 subunits) triangulation are formed. The diameter of the capsids across the tips of the spikes is 36 nm for T = 4 capsids and 32 nm for T = 3 capsids (Crowther et al. 1994). Thus T = 3 capsids provide about 70% of the volume of T = 4 capsids. The asymmetric units of T = 4 and T = 3 capsids both contain

two types of HBc dimers with different spatial surroundings (Fig. 14.4): One type of dimer is located between a fivefold and a local sixfold symmetry axis (blue, cyan in Fig. 14.4) and the other is placed between two local sixfold symmetry axes (red, yellow in Fig. 14.4). In T = 4 capsids the dimer axes of all dimers are located at local twofold symmetry axes, meaning that these dimers do not have perfect twofold symmetry and the individual monomers deviate slightly in their structures to adapt to the different spatial surroundings. In T = 3 capsids only one of the two dimers is located at a local twofold symmetry axes whereas the other dimer is located at a strict two-fold axis and contributes only one of its monomers to the asymmetric unit. The dimer at the strict two-fold symmetry axis in the T = 3 particles has perfect two-fold symmetry with both monomers having identical structures.

Within the capsids the hammerhead-shaped dimers are orientated such that the head is part of the continuous protein shell and the handle forms a protruding spike. Thus capsids have either 90 (T = 3 particle) or 120 (T = 4 particle) spikes at their surface. At either side of the spikes are large triangular holes of approximately 2 nm diameter that fenestrate the capsid shell and are large enough to be crossed by small molecules or nucleotides.

In the capsids the hand-like C-terminal regions of the assembly domains, which form the opposite tips of the hammerhead, mediate the inter-dimer contacts. Either

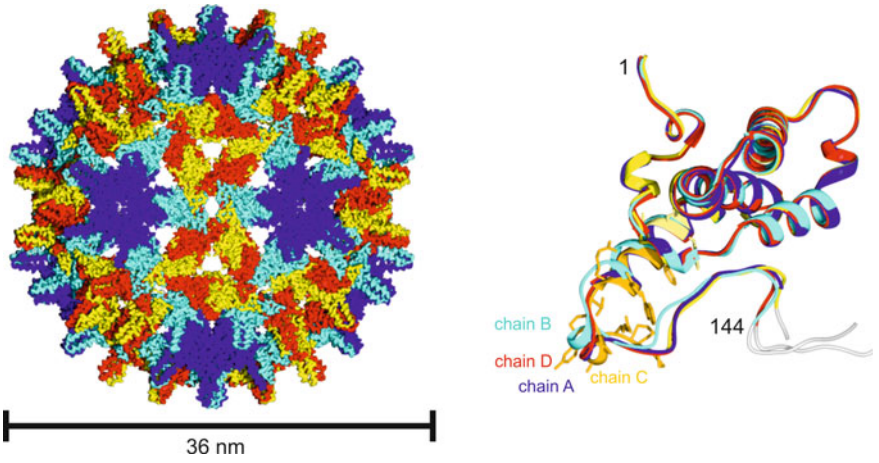


Fig. 14.4 Hbc in the asymmetric unit. Left: Surface representation of a T = 4 capsid [EMDB 0271, (Böttcher and Nassal 2018)]. The surface is coloured according to the spatially different positions of the monomers. One type of dimer (blue, cyan) is located between a fivefold and a twofold symmetry axis. The other type of dimer (red, yellow) is located between two twofold symmetry axes. The maximal diameter of T = 4 capsids measured across the tips of the spikes is 36 nm. Right: The 4 subunits of the asymmetric unit [pdb 6HTX, (Böttcher and Nassal 2018)] superposed. The assembly domain is shown in colour corresponding to the spatial surroundings as indicated on the left. The resolved linker residues are shown in white. The structures deviate in the hand region to adapt to the different spatial surroundings around the five-fold and local six-fold symmetry axes. The most adaptation-affected side chains were identified by NMR and group into a bending region (light yellow) and a dynamic region (dark yellow) (Lecoq et al. 2018) and are indicated for chain C

five or six of these hand-like regions assemble around strict five-fold-symmetry axes and the local six-fold symmetry axes respectively. These different spatial environments require slight adaptations in the structure of the protein (Fig. 14.4) that manifests as a swivel of the hand region relative to the spike with Gly111 as a hinge. Solid State NMR confirms that the hand region comprises a dynamic region between residues 121 and 135 and a bending region between residues 117 and 120 that differ in their chemical shifts between the 4 monomers in the asymmetric unit (Lecoq et al. 2018).

T = 4 capsids are more common than T = 3 capsids and account for more than 90% of capsids in virions (Roseman et al. 2005; Seitz et al. 2007; Dryden et al. 2006). Early experiments with HBc expressed in *E. coli* suggested that the CTD and especially the linker region between assembly domain and ARD (Watts et al. 2002) is decisive for whether T = 3 or T = 4 capsids are formed. In addition full length HBc had a strong dimorphism (Wingfield et al. 1995; Crowther et al. 1994) whereas HBc truncated at position 149 formed mainly T = 4 particles (Zlotnick et al. 1996). Later experiments with optimized, heterologous expression systems did not confirm the earlier described dimorphism of full length wt-HBc (e.g. Böttcher and Nassal 2018.) suggesting that higher dimer concentrations are another decisive factor that controls the speed of capsid assembly (Lutomski et al. 2017) as well as the ratio between T = 3 and T = 4 particles. (Harms et al. 2015). The critical dimer-concentration of truncated $HBc_{1–149}$ for capsid formation is around 0.5 μM and is the same for T = 3 and T = 4 particles (Harms et al. 2015).

Important Sites of the Assembly Domain

HBc is densely packed with functionally important sites. Some of them map to the ordered assembly domain while others are located in the disordered CTD, which probably changes its accessibility during viral maturation. The assembly domain harbours sites that are important for capsid assembly, immunogenicity, formation of virions and secretion (Fig. 14.5). Many functionally important sites map to the inner dimer interface such as the tips of the spikes and the surroundings of a hydrophobic pocket inside the spikes; others are located at the inter-dimer interface like a hydrophobic Heteroaryldihydropyrimidine pocket between the hand region and fulcrum (Venkatakrishnan et al. 2016).

In addition individual, exposed residues of the hand region (R127, I126, F122, A137, N136, and I139) at the outer surface of the spike (L60, L95, and K96) and at the fulcrum (S17, F18) (Fig. 14.5) are essential for the secretion of enveloped virions with a mature genome (Ponsel and Bruss 2003). Even conservative mutations of these key residues abrogate secretion of virions with mature genome (Ponsel and Bruss 2003) but not that of empty, enveloped capsids (Ning et al. 2018).

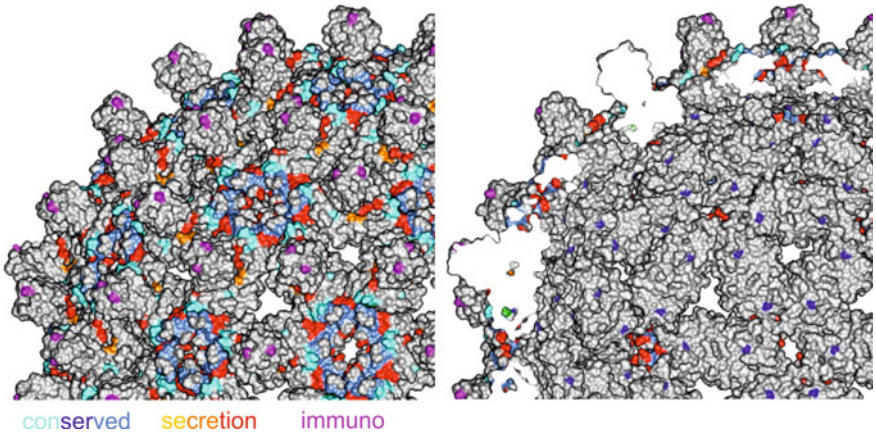


Fig. 14.5 Part of a capsid in surface representation coloured with important sites: evolutionary conserved residues are coloured in blue [motif I light blue, motif II: dark blue, motif III medium blue (Dill et al. 2016)]; residues essential for secretion (Ponsel and Bruss 2003) are coloured in red, mutations causing premature envelopment are shown in yellow and those leading to low level secretion phenotypes in orange (Le Pogam et al. 2000). Magenta highlights residue 80 at the tips of the spikes in the centre of the immune dominant region (Salfeld et al. 1989; Koschel et al. 2000). Green marks sites hat interfere with capsid formation but not with dimer formation. These sites are mainly buried in the protein shell and not surface exposed. The left panel shows the outer surface, to which many important sites map. Interestingly, the upper part of the spikes only contain the immune dominant loop but are otherwise neither conserved nor contain residues that are important for secretion. The right panel shows the inner surface, which only displays 2 important residues: Gly 111 in blue in the conserved motif II and Asn 136 in red being essential for secretion

The Tips of the Spikes

The tips of the spikes are the most exposed site of the capsids and form the major immuno-dominant loop (Fig. 14.5) and probably the binding platform for the envelope. The tips of the spikes consists of the tight loops between helix III and IV containing residues 77–81. Each monomer contributes one of these loops to the tips. This generates a characteristic cleft between the two loops, which could act as potential binding site for other factors. The widths and length of the cleft is only large enough to support direct contact to 2–3 amino acids. The tip of the spikes contains 4 negatively charged residues which accumulate in the cleft providing an exposed platform for potential electrostatic interactions (Fig. 14.6). Furthermore, the tips together with the upper half of the spikes are more flexible than the rest of the capsid as indicated by the high B-factors and the low local resolution in this region of the capsids (Böttcher and Nassal 2018). MD simulations (Hadden et al. 2018) suggest that the tips sample a wide conformational space forming a dynamic cleft between the loops. These observations agree with a potential binding site for small binders following an induced fit mechanism.

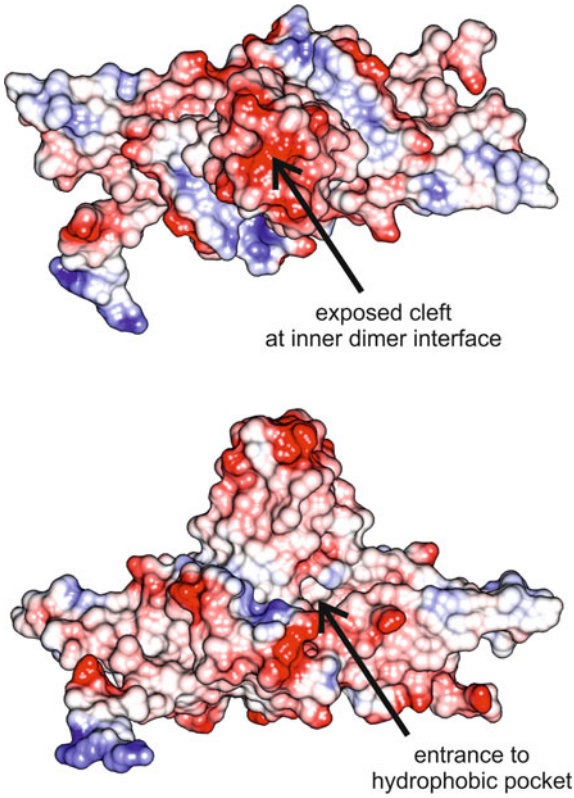


Fig. 14.6 Dimer in surface representation coloured with the relative surface charges. Positive charges are coloured in blue, negative charges in red. The tips of the spikes are the most negatively charged exposed region of the capsids. The fulcrum preceding helix III is also surface exposed and is the most positively charged region of the spike and leads to the entrance of the hydrophobic pocket in the interior of the spikes

The tips of the spikes also harbour the main immunogenic loop, which is centred on residue 80 (Salfeld et al. 1989; Conway et al. 1998; Belnap et al. 2003) at the start of helix IV. Deletion of residue 80 (Fig. 14.5, magenta) abolishes the reaction with human anti-HBc positive serum (Koschel et al. 2000) confirming the location of the immunogenic loop close to the tips of the spikes. Some antibodies show a strong preference for certain quasi-equivalent sites over others (Wu et al. 2013; Conway et al. 2003; Harris et al. 2006) demonstrating that the conformational differences between the four monomers in the asymmetric unit of the capsid can be recognized by antibodies although the structural differences of the four monomers are very minor at this position (Fig. 14.5). In certain complications of Hepatitis B virus infection that lead to acute liver failure, a strong B-cell response causes massive production of IgG against HBc (Farci et al. 2010). Antibodies from such patients bind towards the side of the spikes peripheral to the major immune dominant loop and show a strong conformational preference for some of the monomers over others (Wu et al. 2013). Typically only one monomer per dimer has a bound antibody and within the spatial context of the asymmetric unit it is always the same.

The tips of the spikes are also the contact sites with the envelope in virions (Dryden et al. 2006; Seitz et al. 2007) and bind to fragments of PreS and S-HBs (Muhamad

et al. 2015). NMR-studies suggest that the S-HBs fragment forms a short helical segment and the PreS fragment a tight loop, which both bind with one face to the cleft at the tips of the spikes (Muhamad et al. 2015). Peptides that interfere with binding of L-HBs to the capsids (Dyson and Murray 1995) also attach to the tips of the spikes at the cleft at the tips of the spikes (Böttcher et al. 1998; Freund et al. 2008). Contacts between capsid and envelope appear variable from varying directions (Seitz et al. 2007). Although these observations strongly suggest that the tips of the spikes are the binding sites between capsid and envelope, none of the residues at the tips of the spikes appears to be of importance for envelopment, virion formation or secretion according to extensive mutational screens (Koschel et al. 2000; Ponsel and Bruss 2003). This apparent contradiction is not yet fully resolved and might hint to an unspecific electrostatic interaction between HBc and the envelope and could reflect the different roles of S-HBs and L-HBs in envelopment with PreS probably mediating the specific contact away from the tips of the spikes (Ponsel and Bruss 2003; Pastor et al. 2019).

The Hydrophobic Pocket Inside the Spikes

The spikes accommodate a hydrophobic pocket (Fig. 14.6, lower panel, arrow) in the centre of the spikes, which harbours Cys61 and F/I97. In capsids derived from patients' serum (Roseman et al. 2005) and in some samples of capsids derived from HBc over-expressed in *E. coli* (Böttcher and Nassal 2018) the pocket contains an unknown factor. Its small size is consistent with a small molecule or an individual amino-acid side chain rather than with a folded motif of a potential protein binding partner. The pocket is of interest because it is lined by numerous residues that have been implicated in either low or premature secretion phenotypes. In particular residue 97 is implicated in premature secretion. In wt-Hbc this position is typically occupied by a large hydrophobic residue, such as phenylalanine (e.g. strains ayw and adyw) or isoleucine (e.g. strains adr and adw); however, in some naturally occurring mutants, these bulky residues are replaced by a smaller leucine. These mutants have a premature secretion phenotype such that capsids containing ssDNA are enveloped (Yuan et al. 1999a, b).

Structural differences between capsids of wt-HBc and of F97L-mutant are subtle and impact the side chain orientation of residues 97 in helix IV and the disulphide-bridge at the adjacent helix III, which is always reduced in 97L but not in 97F. Both changes together enlarge the hydrophobic pocket (Böttcher and Nassal 2018) such that the pocket factor can penetrate deeper into the spike. The entrance of the pocket is also lined by P5 and L60 (Fig. 14.5, orange) which if mutated to P5T or to L60V are implicated with low level secretion phenotypes and are naturally occurring variants of the Hepatitis B Virus (Le Pogam et al. 2000). Residue 5 is close to the N-terminus of HBc, which is largely disordered in freshly prepared, phosphorylated capsids and folds slowly over a couple of days (Heger-Stevic et al. 2018b). This slow maturation in vitro might hint at a cis–trans isomerization of the proline which

is generally catalysed by peptidyl-prolyl-cis-trans-isomerases in vivo. Mutations of the proline are likely to abolish the structural maturation and thus interfere with processes that sense a folded N-terminal region. In vivo experiments in Huh cells show that mutating any of the pocket delineating residues L60, L95, K96 and I126 to smaller alanines abrogates binding of the L-HBs to HBc (Pastor et al. 2019). All these observations suggest that the pocket could be involved in binding PreS but is generally too small to accommodate parts of the PreS-sequence directly.

The Hydrophobic Pocket at the Inter Dimer Interface

HBc-dimers interact around the five-fold and local six-fold symmetry axes by packing the hand region of one dimer against the fulcrum and the hand region of the adjacent dimer. This inter-dimer contact forms a hydrophobic pocket between two dimers to which assembly modulators bind (Venkatakrishnan et al. 2016; Schlicksup et al. 2018). These assembly modulators fall into two categories: The heteroaryldihydropyrimidine-derivates (HAP) (Stray et al. 2005) and the phenylpropenamide-derivates (Katen et al. 2010). While HAP-derivates lead to aberrant structures (Stray et al. 2005) phenylpropenamides accelerate capsid assembly and maintain the capsid structure (Katen et al. 2010). Both types of modulators have antiviral potential (for review see Zlotnick et al. 2015) but have a different mode of action. HAPs probably act by accumulating hexamers of dimers as intermediates (Stray et al. 2005). These hexamers are well stabilized and can lead to tube-like assemblies with hexamers of dimers as main building block (Liu et al. 2017). In contrast phenylpropenamides lead to accelerated formation of intact capsids that then progress into empty enveloped virus-like particles (Feld et al. 2007). HAP and Phenylpropenamides both intercalate between helix V of the hand region in one dimer and helix IV of the spike below the fulcrum of the other dimer (Schlicksup et al. 2018; Katen et al. 2013). The changes in the individual asymmetric units are small but lead to large range distortions that accumulate in the capsid structures (Katen et al. 2013).

Most of the structural binding studies of these modulators have used C-terminally truncated $HBc_{1\text{-}149}$-variants. The truncation removes the ARD of the CTD and abolishes unspecific interactions with RNA. The resulting capsids are more homogeneous and therefore are better amenable to high-resolution structure determination. However, in capsids C-terminally truncated HBc and full length HBc have a somewhat different orientation of the C-terminal tail of the assembly domain (Yu et al. 2013), which impacts the inter-dimeric hydrophobic pocket. Superposing structures of truncated HBc variants with bound modulator (Schlicksup et al. 2018; Venkatakrishnan et al. 2016; Katen et al. 2013) with full length HBc-capsids suggests that the modulators also affect residues 151–152 of the adjacent subunit (Böttcher and Nassal 2018) making an even more extended contact at the interdimer interface than suggested by the studies of the truncated $HBc_{1–149}$. It is likely that part of the mode of action of the assembly modulators is in co-ordinating the CTD and accelerating the

interaction of these only weakly ordered residues to form stable capsids. Recently, more substance classes were discovered (Huber et al. 2019; Corcuera et al. 2018) which also target the same hydrophobic pocket and modulate capsid assembly. These modulators provide a versatile toolbox for interfering with viral maturation and thus for the development of new antiviral strategies.

The C-Terminal Domain (CTD)

The CTD downstream of residue 144 is largely disordered. It can be divided into a linker region and the arginine rich domain (ARD). Parts of the linker region have been resolved as weak density in high resolution EM-maps of empty, phosphorylated capsids (Böttcher and Nassal 2018). This linker region folds into a kinked stretch that resembles the loop and stretch of the hand region. The linker packs against the last resolved residues of the assembly domain in the adjacent dimer and contributes to the inter-dimer contact. This involvement in inter-dimer contacts probably explains its impact on the morphology of capsids (Zlotnick et al. 1996).

The ARD downstream of the linker has many positive charges located at 16 arginines and interacts randomly with host-RNA. These arginines are interspersed with six serines, which can be phosphorylated and which are phosphorylated in HBc-dimers prior to capsid assembly (Zhao et al. 2018). There is evidence that the dephosphorylation of HBc is concomitant with the encapsidation of the pregenomic RNA (Zhao et al. 2018). If HBc is phosphorylated in vitro, capsids no longer incorporate random host RNA and remain empty (Heger-Stevic et al. 2018b). Similarly, the packaging of host RNA is reduced by phosphorylation mimics in which three of the serines are replaced by negatively charged glutamates (Porterfield et al. 2010). Thus, stable incorporation of the genome probably requires a fine-tuned charge balance between negative charges contributed by the genome and the phosphorylated serines and the positive charges contributed by the arginines (Le Pogam et al. 2005). Phosphorylation of the CTD has no impact on the structure of the assembly domain (Böttcher and Nassal 2018; Heger-Stevic et al. 2018b). Therefore, a conformational change of the capsid cannot act as a maturation signal that directs the capsid either towards envelopment or to the nucleus. However, differences in the accessibility of the CTD in phosphorylated and unphosphorylated capsids as well as capsids of phosphorylation mimics are evident by a changed accessibility to proteases (Heger-Stevic et al. 2018b; Selzer et al. 2015). Such conformational accessibility-switch is an attractive mode for guiding maturation as it either masks or unmasks the nuclear localization signal. However, it is difficult to understand what the actual driver of this maturation is. Some experiments compare the accessibility of the CTD to trypsin digestion in empty capsids that have been generated from HBc dimers in the absence of RNA (Selzer et al. 2015), whereas other experiments compare capsids that self-assemble in *E. coli* (Heger-Stevic et al. 2018b) and balance their charge by either picking up RNA (unphosphorylated) or remaining empty (phosphorylated). Accordingly, the observations differ and show different degrees of protection of the CTD. Unfortunately,

protease protection assays cannot distinguish whether accessibility has changed due to a partial disintegration of the capsids or a genuine conformational switch. Studies on nucleocapsids with mature genome suggest that the maturation leads to a destabilization of the capsids probably with partial disintegration (Cui et al. 2013), which could explain changes in accessibility to proteases. In contrast, structural studies, and especially those aiming at high-resolution select for the most intact particles and discard information from a large percentage of particles, which are structurally more variable or damaged but could account for the observed changed accessibility.

Assuming a conformational switch that makes the CTD accessible to the outside without damaging the gross architecture of the capsids raises the question where the CTDs emerge from the capsid. Electron microscopic studies of capsids with Import in ß bound to the CTD suggests that the CTD egresses through the holes at the local six-fold axes (Chen et al. 2016). However, these holes are narrow and cannot easily widen to accommodate an extended peptide stretch without a simultaneous rearrangement of the adjacent hand regions. Far more attractive are the large holes at the base of the spikes. With a diameter of some 2 nm, they are large enough to support the exit of one or two extended stretches of the CTD. In agreement with this model are observations on phosphorylated HBc capsids, which show the last resolved residues of the assembly domain extending towards the large holes adjacent to the base of the spikes (Heger-Stevic et al. 2018b; Böttcher and Nassal 2018). Furthermore, in the phosphorylated premature envelopment mutant F97L the three most C-terminal residues of the CTD (181–183) are located at the base of the spikes directly adjacent to the large holes (Böttcher and Nassal 2018). This locates the C-termini of HBc at the capsid interior at a favourable position for exposure through the adjacent large holes without the need of capsid reorganization in the assembly domain.

Evolution of Capsids

Hepadnaviridae and the related Nackednaviridae are found in almost all vertebrates, including fish (Hahn et al. 2015), amphibians (Dill et al. 2016), birds and mammals. Phylogenetic analysis suggests that both families emerged from a common ancestor some 430 Mya (Lauber et al. 2017). While the non-enveloped Nackednaviridae are only reported in fish, the enveloped Hepadnaviridae are found in mammals (Orthohepadnavirusses), birds (Avihepatovirusses), reptiles, amphibian (Herpetohepadnavirusses) and in some fish (Meta- and Parahepdanaviruses) (Lauber et al. 2017; Schaefer 2007). The non-enveloped Nackednaviridae have a capsid that is structurally similar to the human HBc capsids (Lauber et al. 2017) but probably evolved much earlier. Phylogenetic comparison identify three conserved motifs in the capsid proteins (Dill et al. 2016), which include the surface-exposed side of the hand region, parts of the fulcrum and the tight kink between helix IV and IV connecting the spike domain to the hand region (Fig. 14.5, blue). All three motifs are in close proximity to the hydrophobic pocket at the inter dimer interface, suggesting that important sites for capsid formation are conserved over geological ages. In contrast, residues at

the inner-dimer interface and at the tips of the spikes are not conserved. While most capsid proteins are relatively small with some 180+ residues the capsid proteins of Avihepadnaviruses in birds and Herpetohepadnaviruses in reptiles and amphibians have much larger capsid proteins of some 260+ residues. The most remarkable difference of these larger variants compared to the other capsid proteins are extensions of some 40–50 residues replacing the tips of the spikes (Nassal et al. 2007). This demonstrates the structural variability of the spike region. Recent unpublished data shows that in duck Hepatitis B capsid protein (DHBc) this extension domain folds separately and much more slowly than the rest of the capsid protein (Makbul et al. 2020), The extension domain packs to the side of the spikes as already suggested by the earlier discovered broad spikes of DHBc capsids (Kenney et al. 1995) and contributes almost 50% of the area of the inner-dimer interface. Thus it adds an important contribution to capsid stability (Makbul et al. 2020). The inner dimer contact within the core spike of DHBc is also changed in comparison to HBc with an altered twist of the helices III and IV and a stabilization of the contact by stacking of aromatic residues rather than an inner-dimer di-sulphide bridge (Makbul et al. 2020). These observations further highlight the fact that the inner-dimer interface in the related core proteins is much less conserved than the inter-dimer packing.

HBc Capsids as Tool in Biotechnology

The ability of HBc to self-assemble into capsids has been fascinating for many researchers and was quickly developed into a tool to display other proteins or protein fragments at the surface of capsids (Schodel et al. 1992; Borisova et al. 1996; Ulrich et al. 1998). Large insertions are tolerated at the tips of the spikes without abolishing particle formation. Depending on the size of the inserted protein these particles have some 40–60 nm diameter and carry up to 240 foreign protein fragments at their surface. This spatially condensed display of protruding epitopes is conceptually interesting for the generation of vaccines that trigger a strong immune response. HBc is a particularly attractive carrier, as it is not infectious on its own, triggers a strong B cell, T cell and CTL response (Milich et al. 1995) and is directly displayed in B cells in mice (Milich et al. 1997). Systematic structural screening has shown that the tips of the spikes tolerate the insertion of whole proteins such as GFP (Kratz et al. 1999; Böttcher et al. 2006) or the dimeric outer surface lipoprotein C (OspC) of the Lyme disease (Skamel et al. 2006). However, insertions into HBc at the tips of the spikes requires that the C- and N-terminus of the inserted fragment are in close proximity and that spatial clashes of the two displayed proteins on the same spike do not interfere with the independent folding of HBc and the inserted protein. Some residual conformational stress at the tips of the spikes is generally well tolerated by the capsids but is propagated through the capsid scaffold and leads to distortions of the capsid that are hinged at Gly 111 at the joint between helix IV in the spike and helix V in the hand-region (Böttcher et al. 2006). This type of conformational stress occurs during protein folding and capsid formation. It is also possible

to induce conformational stress post capsid formation. For this an acidic insertion domain is introduced which forms a coiled coil helix with a complementary basic peptide (peptide velcro) but is disordered without the binding partner (O'Shea et al. 1993). In the absence of the complimentary peptide, capsids have the expected structure of wt-HBc capsids. The large acidic insertions are not resolved by electron cryo-microscopy, which is expected for an unfolded insert. Upon addition of the complimentary basic peptide, capsids completely disintegrate (Böttcher et al. 2006). This enables a triggered opening of an artificial HBc-based nano-container.

Spatial constraints restrict which proteins and protein fragments can be inserted to the tips of the spikes. These constraints are partly eliminated by using very long flexible linkers, which give the foreign fragment the freedom to fold independently of the icosahedral capsid scaffold. Similarly, using mosaic particles of wt-HBc and chimeric HBc (Vogel et al. 2005) with an inserted foreign protein (fragment) relieve the spatial constraints by reducing the number of surface-exposed foreign epitopes. However, insertions of proteins in which C- and N-terminus are distant remain problematic. Here a different strategy which uses "split cores" was designed (Walker et al. 2011). In these constructs HBc_{1-149} is split at the tips of the spikes between helices III and IV into an N-terminal and a C-terminal protein. These two split-core proteins still assemble into icosahedral capsids and also do so when foreign proteins are inserted at the tips of the spikes. This strategy by Walker and co-workers allows the expression of the foreign proteins at the C-or N-terminal end of the split-core rather than in a conformationally more constraint loop and has been successfully used for more problematic inserts.

Acknowledgements BB is supported by the German Research Foundation (DFG, BO1150/15-1).

References

Alexander CG, Jurgens MC, Shepherd DA, Freund SM, Ashcroft AE, Ferguson N (2013) Thermodynamic origins of protein folding, allostery, and capsid formation in the human hepatitis B virus core protein. Proc Natl Acad Sci USA 110(30):E2782-2791. https://doi.org/10.1073/pnas.1308846110[pii]

Alter MJ (2003) Epidemiology of hepatitis B in Europe and worldwide. J Hepatol 39(Suppl 1):S64-69

Belnap DM, Watts NR, Conway JF, Cheng N, Stahl SJ, Wingfield PT, Steven AC (2003) Diversity of core antigen epitopes of hepatitis B virus. Proc Natl Acad Sci USA 100(19):10884–10889

Borisova G, Borschukova Wanst O, Mezule G, Skrastina D, Petrovskis I, Dislers A, Pumpens P, Grens E (1996) Spatial structure and insertion capacity of immunodominant region of hepatitis B core antigen. Intervirology 39(1–2):16–22. https://doi.org/10.1159/000150470

Böttcher B, Nassal M (2018) Structure of mutant hepatitis B core protein capsids with premature secretion phenotype. J Mol Biol 430:4941–4954. https://doi.org/10.1016/j.jmb.2018.10.018

Bottcher B, Wynne SA, Crowther RA (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386(6620):88–91. https://doi.org/10.1038/386088a0

Böttcher B, Vogel M, Ploss M, Nassal M (2006) High plasticity of the hepatitis B virus capsid revealed by conformational stress. J Mol Biol 356(3):812–822

Böttcher B, Tsuji N, Takahashi H, Dyson MR, Zhao S, Crowther RA, Murray K (1998) Peptides that block hepatitis B virus assembly: analysis by cryomicroscopy, mutagenesis and transfection. EMBO J 17(23):6839–6845

Bourne CR, Finn MG, Zlotnick A (2006) Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. J Virol 80(22):11055–11061. JVI.00933-06 [pii] https://doi.org/10.1128/JVI.00933-06

Bruss V, Lu X, Thomssen R, Gerlich WH (1994) Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. EMBO J 13(10):2273–2279

Chen C, Wang JC, Pierson EE, Keifer DZ, Delaleau M, Gallucci L, Cazenave C, Kann M, Jarrold MF, Zlotnick A (2016) Importin beta can bind hepatitis B virus core protein and empty core-like particles and induce structural changes. PLoS Pathog 12(8):e1005802. https://doi.org/10.1371/journal.ppat.1005802

Conway JF, Chenng N, Zlotnick A, Wingfield PT, Stahl SJ, Steven AC (1997) Visualization of a4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. Nature 368(6):91–94

Conway JF, Watts NR, Belnap DM, Cheng N, Stahl SJ, Wingfield PT, Steven AC (2003) Characterization of a conformational epitope on hepatitis B virus core antigen and quasiequivalent variations in antibody binding. J Virol 77(11):6466–6473

Conway JF, Cheng N, Zlotnick A, Stahl SJ, Wingfield PT, Belnap DM, Kanngiesser U, Noah M, Steven AC (1998) Hepatitis B virus capsid: localization of the putative immunodominant loop (residues 78 to 83) on the capsid surface, and implications for the distinction between c and e-antigens. J Mol Biol 279(5):1111–1121

Corcuera A, Stolle K, Hillmer S, Seitz S, Lee JY, Bartenschlager R, Birkmann A, Urban A (2018) Novel non-heteroarylpyrimidine (HAP) capsid assembly modifiers have a different mode of action from HAPs in vitro. Antiviral Res 158:135–142. https://doi.org/10.1016/j.antiviral.2018.07.011

Crowther RA, Kiselev NA, Bottcher B, Berriman JA, Borisova GP, Ose V, Pumpens P (1994) Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell 77:943–950

Cui X, Ludgate L, Ning X, Hu J (2013) Maturation-associated destabilization of hepatitis B virus nucleocapsid. J Virol 87(21):11494–11503. https://doi.org/10.1128/JVI.01912-13

Dill JA, Camus AC, Leary JH, Di Giallonardo F, Holmes EC, Ng TF (2016) Distinct viral lineages from fish and amphibians reveal the complex evolutionary history of hepadnaviruses. J Virol 90(17):7920–7933. https://doi.org/10.1128/JVI.00832-16

Dryden KA, Wieland SF, Whitten-Bauer C, Gerin JL, Chisari FV, Yeager M (2006) Native hepatitis B virions and capsids visualized by electron cryomicroscopy. Mol Cell 22(6):843–850

Dyson MR, Murray K (1995) Selection of peptide inhibitors of interactions involved in complex protein assemblies: association of the core and surface antigens of hepatitis B virus. Proc Natl Acad Sci USA 92:2194–2198

Farci P, Diaz G, Chen Z, Govindarajan S, Tice A, Agulto L, Pittaluga S, Boon D, Yu C, Engle RE, Haas M, Simon R, Purcell RH, Zamboni F (2010) B cell gene signature with massive intrahepatic production of antibodies to hepatitis B core antigen in hepatitis B virus-associated acute liver failure. Proc Natl Acad Sci USA 107(19):8766–8771. https://doi.org/10.1073/pnas.1003854107

Feld JJ, Colledge D, Sozzi V, Edwards R, Littlejohn M, Locarnini SA (2007) The phenylpropenamide derivative AT-130 blocks HBV replication at the level of viral RNA packaging. Antiviral Res 76(2):168–177. https://doi.org/10.1016/j.antiviral.2007.06.014

Freund SM, Johnson CM, Jaulent AM, Ferguson N (2008) Moving towards high-resolution descriptions of the molecular interactions and structural rearrangements of the human hepatitis B core protein. J Mol Biol 384(5):1301–1313. S0022-2836(08)01292-8 [pii] https://doi.org/10.1016/j.jmb.2008.10.020

Hadden JA, Perilla JR, Schlicksup CJ, Venkatakrishnan B, Zlotnick A, Schulten K (2018) All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. eLife 7. https://doi.org/10.7554/eLife.32478

Hahn CM, Iwanowicz LR, Cornman RS, Conway CM, Winton JR, Blazer VS (2015) Characterization of a novel hepadnavirus in the white sucker (*Catostomus commersonii*) from the Great Lakes Region of the United States. J Virol 89(23):11801–11811. https://doi.org/10.1128/JVI.01278-15

Harms ZD, Selzer L, Zlotnick A, Jacobson SC (2015) Monitoring assembly of virus capsids with nanofluidic devices. ACS Nano 9(9):9087–9096. https://doi.org/10.1021/acsnano.5b03231

Harris A, Belnap DM, Watts NR, Conway JF, Cheng N, Stahl SJ, Vethanayagam JG, Wingfield PT, Steven AC (2006) Epitope diversity of hepatitis B virus capsids: quasi-equivalent variations in spike epitopes and binding of different antibodies to the same epitope. J Mol Biol 355(3):562–576

Heger-Stevic J, Kolb P, Walker A, Nassal M (2018a) Displaying whole-chain proteins on hepatitis B virus capsid-like particles. Methods Mol Biol 1776:503–531. https://doi.org/10.1007/978-1-4939-7808-3_33

Heger-Stevic J, Zimmermann P, Lecoq L, Bottcher B, Nassal M (2018b) Hepatitis B virus core protein phosphorylation: identification of the SRPK1 target sites and impact of their occupancy on RNA binding and capsid structure. PLoS Pathog 14(12):e1007488. https://doi.org/10.1371/journal.ppat.1007488

Huber AD, Pineda DL, Liu D, Boschert KN, Gres AT, Wolf JJ, Coonrod EM, Tang J, Laughlin TG, Yang Q, Puray-Chavez MN, Ji J, Singh K, Kirby KA, Wang Z, Sarafianos SG (2019) Novel hepatitis B virus capsid-targeting antiviral that aggregates core particles and inhibits nuclear entry of viral cores. ACS Infect Dis 5(5):750–758. https://doi.org/10.1021/acsinfecdis.8b00235

Katen SP, Chirapu SR, Finn MG, Zlotnick A (2010) Trapping of hepatitis B virus capsid assembly intermediates by phenylpropenamide assembly accelerators. ACS Chem Biol 5(12):1125–1136. https://doi.org/10.1021/cb100275b

Katen SP, Tan Z, Chirapu SR, Finn MG, Zlotnick A (2013) Assembly-directed antivirals differentially bind quasiequivalent pockets to modify hepatitis B virus capsid tertiary and quaternary structure. Structure 21(8):1406–1416. https://doi.org/10.1016/j.str.2013.06.013S0969-2126(13)00214-1[pii]

Kenney B, Nassal F (1995) Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. Structure 3:1009–1019

Klumpp K, Lam AM, Lukacs C, Vogel R, Ren S, Espiritu C, Baydo R, Atkins K, Abendroth J, Liao G, Efimov A, Hartman G, Flores OA (2015) High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. Proc Natl Acad Sci USA 112(49):15196–15201. https://doi.org/10.1073/pnas.1513803112

Koschel M, Oed D, Gerelsaikhan T, Thomssen R, Bruss V (2000) Hepatitis B virus core gene mutations which block nucleocapsid envelopment. J Virol 74(1):1–7

Kratz PA, Bottcher B, Nassal M (1999) Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. Proc Natl Acad Sci USA 96(5):1915–1920

Lauber C, Seitz S, Mattei S, Suh A, Beck J, Herstein J, Borold J, Salzburger W, Kaderali L, Briggs JAG, Bartenschlager R (2017) Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. Cell Host Microbe 22(3):387–399 e386. https://doi.org/10.1016/j.chom.2017.07.019

Lecoq L, Wang S, Wiegand T, Bressanelli S, Nassal M, Meier BH, Bockmann A (2018) Localizing conformational hinges by NMR: where do hepatitis B virus core proteins adapt for capsid assembly? ChemPhysChem 19(11):1336–1340. https://doi.org/10.1002/cphc.201800211

Le Pogam S, Chua PK, Newman M, Shih C (2005) Exposure of RNA templates and encapsidation of spliced viral RNA are influenced by the arginine-rich domain of human hepatitis B virus core antigen (HBcAg 165-173). J Virol 79(3):1871–1887. 79/3/1871 [pii] https://doi.org/10.1128/JVI.79.3.1871-1887.2005

Le Pogam S, Yuan TT, Sahu GK, Chatterjee S, Shih C (2000) Low-level secretion of human hepatitis B virus virions caused by two independent, naturally occurring mutations (P5T and L60V) in the capsid protein. J Virol 74(19):9099–9105

Liu C, Fan G, Wang Z, Chen HS, Yin CC (2017) Allosteric conformational changes of human HBV core protein transform its assembly. Sci Rep 7(1):1404. https://doi.org/10.1038/s41598-017-01568-9

Lutomski CA, Lyktey NA, Zhao Z, Pierson EE, Zlotnick A, Jarrold MF (2017) Hepatitis B virus capsid completion occurs through error correction. J Am Chem Soc 139(46):16932–16938. https://doi.org/10.1021/jacs.7b09932

Makbul C, Nassal M, Böttcher B (2020) Slowly folding surface extension in duck hepatitis B virus capsids governs its stability. eLife 9:e57277. https://doi.org/10.7554/eLife.57277

Milich DR, Peterson DL, Zheng J, Hughes JL, Wirtz R, Schodel F (1995) The hepatitis nucleocapsid as a vaccine carrier moiety. Ann N Y Acad Sci 754:187–201. https://doi.org/10.1111/j.1749-6632.1995.tb44451.x

Milich DR, Chen M, Schodel F, Peterson DL, Jones JE, Hughes JL (1997) Role of B cells in antigen presentation of the hepatitis B core. Proc Natl Acad Sci USA 94(26):14648–14653. https://doi.org/10.1073/pnas.94.26.14648

Muhamad A, Ho KL, Rahman MBA, Tejo BA, Uhrin D, Tan WS (2015) Hepatitis B virus peptide inhibitors: solution structures and interactions with the viral capsid. Org Biomol Chem 13(28):7780–7789. https://doi.org/10.1039/c5ob00449g

Nassal M (1992) Conserved cysteines of the hepatitis B virus core protein are not required for assembly of replication-competent core particles nor for their envelopment. Virology 190(1):499–505

Nassal M (2008) Hepatitis B viruses: reverse transcription a different way. Virus Res 134(1–2):235–249. https://doi.org/10.1016/j.virusres.2007.12.024

Nassal M, Leifer I, Wingert I, Dallmeier K, Prinz S, Vorreiter J (2007) A structural model for duck hepatitis B virus core protein derived by extensive mutagenesis. J Virol 81(23):13218–13229. JVI.00846-07 [pii] https://doi.org/10.1128/JVI.00846-07

Ning X, Luckenbaugh L, Liu K, Bruss V, Sureau C, Hu J (2018) Common and distinct capsid and surface protein requirements for secretion of complete and genome-free hepatitis B virions. J Virol 92(14). https://doi.org/10.1128/JVI.00272-18

Ning X, Nguyen D, Mentzer L, Adams C, Lee H, Ashley R, Hafenstein S, Hu J (2011) Secretion of genome-free hepatitis B virus–single strand blocking model for virion morphogenesis of para-retrovirus. PLoS Pathog 7(9):e1002255. https://doi.org/10.1371/journal.ppat.1002255

Ning X, Basagoudanavar SH, Liu K, Luckenbaugh L, Wei D, Wang C, Wei B, Zhao Y, Yan T, Delaney W, Hu J (2017) Capsid phosphorylation state and hepadnavirus virion secretion. J Virol 91(9). https://doi.org/10.1128/JVI.00092-17

O'Shea EK, Lumb KJ, Kim PS (1993) Peptide 'Velcro': design of a heterodimeric coiled coil. Curr Biol 3(10):658–667

Packianathan C, Katen SP, Dann CE, 3rd, Zlotnick A (2010) Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. J Virol 84(3):1607–1615. JVI.02033-09 [pii] https://doi.org/10.1128/JVI.02033-09

Pastor F, Herrscher C, Patient R, Eymieux S, Moreau A, Burlaud-Gaillard J, Seigneuret F, de Rocquigny H, Roingeard P, Hourioux C (2019) Direct interaction between the hepatitis B virus core and envelope proteins analyzed in a cellular context. Sci Rep 9(1):16178. https://doi.org/10.1038/s41598-019-52824-z

Ponsel D, Bruss V (2003) Mapping of amino acid side chains on the surface of hepatitis B virus capsids required for envelopment and virion formation. J Virol 77(1):416–422

Porterfield JZ, Dhason MS, Loeb DD, Nassal M, Stray SJ, Zlotnick A (2010) Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. J Virol 84(14):7174–7184. JVI.00586-10 [pii] https://doi.org/10.1128/JVI.00586-10

Qiu Z, Lin X, Zhou M, Liu Y, Zhu W, Chen W, Zhang W, Guo L, Liu H, Wu G, Huang M, Jiang M, Xu Z, Zhou Z, Qin N, Ren S, Qiu H, Zhong S, Zhang Y, Zhang Y, Wu X, Shi L, Shen F, Mao Y, Zhou X, Yang W, Wu JZ, Yang G, Mayweg AV, Shen HC, Tang G (2016) Design and synthesis of orally bioavailable 4-methyl heteroaryldihydropyrimidine based hepatitis B virus (HBV) capsid inhibitors. J Med Chem 59(16):7651–7666. https://doi.org/10.1021/acs.jmedchem.6b00879

Roseman AM, Berriman JA, Wynne SA, Butler PJ, Crowther RA (2005) A structural model for maturation of the hepatitis B virus core. Proc Natl Acad Sci USA 102(44):15821–15826

Salfeld J, Pfaff E, Noah M, Schaller H (1989) Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. J Virol 63(2):798–808

Schaefer S (2007) Hepatitis B virus taxonomy and hepatitis B virus genotypes. World J Gastroenterol 13(1):14–21. https://doi.org/10.3748/wjg.v13.i1.14

Schlicksup CJ, Wang JC, Francis S, Venkatakrishnan B, Turner WW, VanNieuwenhze M, Zlotnick A (2018) Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. eLife 7. https://doi.org/10.7554/eLife.31473

Schodel F, Moriarty AM, Peterson DL, Zheng JA, Hughes JL, Will H, Leturcq DJ, McGee JS, Milich DR (1992) The position of heterologous epitopes inserted in hepatitis B virus core particles determines their immunogenicity. J Virol 66(1):106–114

Seitz S, Urban S, Antoni C, Bottcher B (2007) Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. EMBO J 26(18):4160–4167. 7601841 [pii] https://doi.org/10.1038/sj.emboj.7601841

Selzer L, Katen SP, Zlotnick A (2014) The hepatitis B virus core protein intradimer interface modulates capsid assembly and stability. Biochemistry 53(34):5496–5504. https://doi.org/10.1021/bi500732b

Selzer L, Kant R, Wang JCY, Bothner B, Zlotnick A (2015) Hepatitis B virus core protein phosphorylation sites affect capsid stability and transient exposure of the C-terminal domain. J Bio Chem 290(47):28584-+. https://doi.org/10.1074/jbc.M115.678441

Skamel C, Ploss M, Böttcher B, Stehle T, Wallich R, Simon MM, Nassal M (2006) Hepatitis B virus capsid-like particles can display the complete, dimeric outer surface protein C and stimulate production of protective antibody responses against Borrelia burgdorferi infection. J Biol Chem 281(25):17474–17481

Stray SJ, Bourne CR, Punna S, Lewis WG, Finn MG, Zlotnick A (2005) A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. Proc Natl Acad Sci USA 102(23):8138–8143

Summers J, Mason WS (1982) Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. Cell 29(2):403–415. 0092-8674(82)90157-X [pii]

Ulrich R, Nassal M, Meisel H, Kruger DH (1998) Core particles of hepatitis B virus as carrier for foreign epitopes. Adv Virus Res 50:141–182. https://doi.org/10.1016/s0065-3527(08)60808-8

Venkatakrishnan B, Katen SP, Francis S, Chirapu S, Finn MG, Zlotnick A (2016) Hepatitis B virus capsids have diverse structural responses to small-molecule ligands bound to the heteroaryldihydropyrimidine pocket. J Virol 90(8):3994–4004. https://doi.org/10.1128/JVI.03058-15

Vogel M, Diez M, Eisfeld J, Nassal M (2005) In vitro assembly of mosaic hepatitis B virus capsid-like particles (CLPs): rescue into CLPs of assembly-deficient core protein fusions and FRET-suited CLPs. FEBS Lett 579(23):5211–5216. https://doi.org/10.1016/j.febslet.2005.08.044

Walker A, Skamel C, Nassal M (2011) SplitCore: an exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. Sci Rep 1:5. https://doi.org/10.1038/srep00005

Watts NR, Conway JF, Cheng N, Stahl SJ, Belnap DM, Steven AC, Wingfield PT (2002) The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface. EMBO J 21(5):876–884

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

Wu W, Chen Z, Cheng N, Watts NR, Stahl SJ, Farci P, Purcell RH, Wingfield PT, Steven AC (2013) Specificity of an anti-capsid antibody associated with Hepatitis B Virus-related acute liver failure. J Struct Biol 181(1):53–60. https://doi.org/10.1016/j.jsb.2012.10.004

Wynne SA, Crowther RA, Leslie AG (1999) The crystal structure of the human hepatitis B virus capsid. Mol Cell 3(6):771–780

Yuan TT, Sahu GK, Whitehead WE, Greenberg R, Shih C (1999) The mechanism of an immature secretion phenotype of a highly frequent naturally occurring missense mutation at codon 97 of human hepatitis B virus core antigen. J Virol 73(7):5731–5740

Yuan TT, Tai PC, Shih C (1999) Subtype-independent immature secretion and subtype-dependent replication deficiency of a highly frequent, naturally occurring mutation of human hepatitis B virus core antigen. J Virol 73(12):10122–10128

Yu X, Jin L, Jih J, Shih C, Zhou ZH (2013) 3.5A cryoEM Structure of hepatitis B virus core assembled from full-length core protein. PLoS One 8(9):e69729. https://doi.org/10.1371/journal.pone.0069729 PONE-D-13-04576 [pii]

Zhao Q, Hu Z, Cheng J, Wu S, Luo Y, Chang J, Hu J, Guo JT (2018) Hepatitis B virus core protein dephosphorylation occurs during pregenomic RNA encapsidation. J Virol 92(13). https://doi.org/10.1128/JVI.02139-17

Zhao Z, Wang JC, Gonzalez-Gutierrez G, Venkatakrishnan B, Asor R, Khaykelson D, Raviv U, Zlotnick A (2019) Structural differences between the woodchuck hepatitis virus core protein in the dimer and capsid states are consistent with entropic and conformational regulation of assembly. J Virol 93(14). https://doi.org/10.1128/JVI.00141-19

Zhou Z, Hu T, Zhou X, Wildum S, Garcia-Alcalde F, Xu Z, Wu D, Mao Y, Tian X, Zhou Y, Shen F, Zhang Z, Tang G, Najera I, Yang G, Shen HC, Young JA, Qin N (2017) Heteroaryldihydropyrimidine (HAP) and sulfamoylbenzamide (SBA) inhibit hepatitis B virus replication by different molecular mechanisms. Sci Rep 7:42374. https://doi.org/10.1038/srep42374

Zlotnick A, Venkatakrishnan B, Tan Z, Lewellyn E, Turner W, Francis S (2015) Core protein: a pleiotropic keystone in the HBV lifecycle. Antiviral Res 121:82–93. https://doi.org/10.1016/j.antiviral.2015.06.020

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(23):7412–7421

Chapter 15
Fibrinogen and Fibrin

Rustem I. Litvinov, Marlien Pieters, Zelda de Lange-Loots, and John W. Weisel

Abstract Fibrinogen is a large glycoprotein, synthesized primarily in the liver. With a normal plasma concentration of 1.5–3.5 g/L, fibrinogen is the most abundant blood coagulation factor. The final stage of blood clot formation is the conversion of soluble fibrinogen to insoluble fibrin, the polymeric scaffold for blood clots that stop bleeding (a protective reaction called hemostasis) or obstruct blood vessels (pathological thrombosis). Fibrin is a viscoelastic polymer and the structural and mechanical properties of the fibrin scaffold determine its effectiveness in hemostasis and the development and outcome of thrombotic complications. Fibrin polymerization comprises a number of consecutive reactions, each affecting the ultimate 3D porous network structure. The physical properties of fibrin clots are determined by structural features at the individual fibrin molecule, fibrin fiber, network, and whole clot levels and are among the most important functional characteristics, enabling the blood clot to withstand arterial blood flow, platelet-driven clot contraction, and other dynamic forces. This chapter describes the molecular structure of fibrinogen, the conversion of fibrinogen to fibrin, the mechanical properties of fibrin as well as

R. I. Litvinov
Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, 421 Curie Blvd, BRB II/III, Room 1116, Philadelphia, PA 19104-6058, USA
e-mail: litvinov@pennmedicine.upenn.edu

M. Pieters (✉) · Z. de Lange-Loots
Centre of Excellence for Nutrition, North-West University, Private Bag X6001/Hoffman Street 11, Potchefstroom 2520, South Africa
e-mail: marlien.pieters@nwu.ac.za

Z. de Lange-Loots
e-mail: zelda.delange@nwu.ac.za

Medical Research Council Unit for Hypertension and Cardiovascular Disease, North-West University, Potchefstroom, South Africa

J. W. Weisel
Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, 421 Curie Blvd, BRB II/III, Room 1154, Philadelphia, PA 19104-6058, USA
e-mail: weisel@pennmedicine.upenn.edu

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_15

its structural origins and lastly provides evidence for the role of altered fibrin clot properties in both thrombosis and bleeding.

Keywords Fibrinogen · Fibrin · Viscoelastic polymer · Hemostasis · Thrombosis

Fibrinogen Structure

Quaternary Structure of Fibrinogen

Human fibrinogen is a large 340-kDa oligomeric glycoprotein made up of two symmetrical halves, each containing three polypeptide chains designated Aα, Bβ and γ, with molecular weights of 66,500, 52,000, and 46,500 Da, respectively (Fig. 15.1). The dimeric subunit composition of fibrinogen is presented as $(A\alpha\ B\beta\ \gamma)_2$ to show that fibrinopeptides A (FpA) and B (FpB) are cleaved by thrombin from the N-termini of each of the Aα and Bβ chains, respectively, to form α and β chains in fibrin. The γ chains remain unaffected by thrombin and the formation of monomeric fibrin from fibrinogen can be described as follows: $(A\alpha\ B\beta\ \gamma)_2 \rightarrow (\alpha\ \beta\ \gamma)_2 + 2FpA + 2FpB$. The six subunits are held together by 29 disulfide bonds concentrated in the central nodule of the fibrinogen molecule (Henschen and McDonagh 1986) (Fig. 15.1).

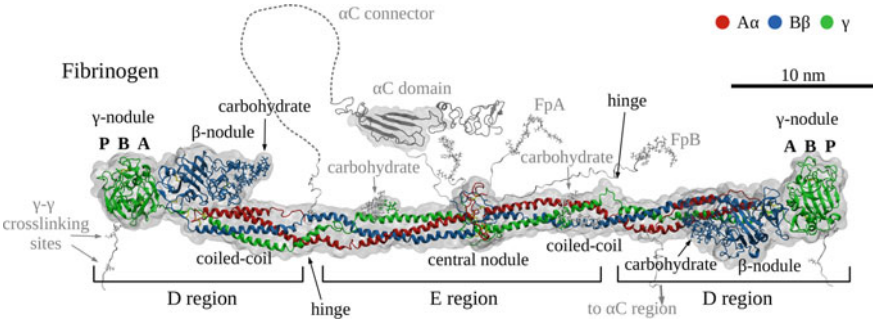


Fig. 15.1 Fibrinogen structure. The atomic resolution structure of about two-thirds of the fibrinogen molecule has been determined by X-ray crystallography (PDB Entry: 3GHG). The folded core of the fibrinogen molecule is shown with addition of portions missing from the crystal structure and reconstructed computationally (in grey), namely: (i) the N-terminal ends of the Aα and Bβ chains with FpA and FpB extending from the central nodule, (ii) the C-terminal portions of the γ-chains containing γ-γ-crosslinking sites, (iii) the αC region composed of the flexible αC connector and compact αC domain (shown on the left halve of the molecule), and (iv) carbohydrate moieties. *P*, *B,* and *A* correspond to the C-terminal A-domain, central B-domain and N-terminal P-domain, respectively, in the γ-nodules. The hinge point is shown in the middle of the α-helical coiled-coil connectors Modified from (Zhmurov et al. 2016)—published with permission from Elsevier

Tertiary Structure of Fibrinogen

The overall shape of fibrinogen molecules studied using X-ray crystallography, transmission electron microscopy, and atomic force microscopy has been described as a rod-like structure of 45 nm long and 2–5 nm thick (Fig. 15.1) (Hall and Slayter 1959; Fowler and Erickson 1979; Williams 1981; Weisel et al. 1985). Each fibrinogen molecule is composed of two lateral D regions and one central E region, each containing a globular part and a portion of α-helical coiled-coils that connect the lateral globules with the central one (Fig. 15.1) (Medved and Weisel 2009). The X-ray crystallography provided atomic resolution structure of the compact core of fibrinogen (Spraggon et al. 1997; Yee et al. 1997; Brown et al. 2000; Madrazo et al. 2001) while about 30% of the mass of the molecule comprising the unstructured and flexible portions was unresolved. The motifs that could not be crystallized and remained uncharacterized structurally at the atomic level are N-terminal residues Aα1-26, Bβ1-57, and γ1-13 and C-terminal residues Aα201-610, Bβ459-461, and γ395-411 (Kollman et al. 2009).

The largest parts of fibrinogen that are missing in the crystal structure are named the αC regions that comprise residues Aα221-610 at the C-terminal parts of the Aα chains. The Aα chain forms a fourth strand at the distal parts of triple α-helical coiled-coils and then extends outward from the bulk of fibrinogen molecule, making the unstructured αC region (Veklich et al. 1993). Each αC region is tethered to the core of the fibrinogen molecule via a flexible connector (residues Aα221-391) linked to the relatively structured but unstable C-terminal domains (residues Aα392-610) that form transitory intramolecular complexes with the central globule (Fig. 15.1) (Weisel and Medved 2001; Litvinov et al. 2007). These 390 residue-long αC regions were visualized in fibrinogen using transmission electron microscopy (Veklich et al. 1993) and atomic force microscopy (Protopopova et al. 2015; Protopopova et al. 2017) as highly flexible appendages. To recover the complete molecular structure of fibrinogen, the crystallographically unresolved parts have been recreated in silico, i.e. computationally (Fig. 15.1) (Zhmurov et al. 2016; Zuev et al. 2017; Zhmurov et al. 2018).

Domain Structure of Fibrinogen

The fibrinogen molecule is organized into structurally and functionally independent units folded into relatively compact domain structures. The nomenclature for fibrinogen and fibrin recommended by the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis (Medved and Weisel 2009) implies that each D or E region contains a number of domains with its own designation, which is different from the name of the region. Based on the X-ray crystallographic data, the central E region is segregated into four domains and each

distal D region is divided into seven structural domains. A triple α-helical *coiled-coil-E domain* containing all three chains is formed next to the *central nodule* in each of two symmetrical halves of the E region. The N-terminal portions of both γ chains connected at the center of the E region form the single asymmetric γ*N-domain*. On the opposite side of the globular portion of the E region, the N-termini of two Aα and two Bβ chains form the *funnel-shaped domain*. In each of the D regions, a triple α-helical *coiled-coil-D domain* is formed next to the globular part by the intertwined portions of all three chains. The C-terminal portion of the β chain is folded into the *β-nodule* (*or β-module*) and the C-terminal portion of the γ chain forms a compact *γ-nodule* (*or γ-module*). Both the β-and γ-nodules each consist of three distinct domains identified in the crystal structure and designated as N-terminal *A-domain*, central *B-domain* and C-terminal *P-domain* roughly shown in Fig. 15.1 (Medved and Weisel 2009). Notably, the structurally distinct domains of a fibrinogen molecule contain functionally important binding sites mediating multiple molecular interactions of fibrinogen.

The Coiled-Coil Connectors

Remarkable and functionally important parts of fibrinogen are the two 17 nm long α-helical coiled-coils that join the central nodule and end globular parts (Fig. 15.1). The α-helical connectors are formed of the intertwined segments of three polypeptide chains, Aα Bβ and γ, each segment containing 111 or 112 amino acid residues. The chains form three right-handed α-helices that wind around each other to form a left-handed supercoil (Cohen and Parry 1990). Importantly, each of the coiled-coils has a fourth helix in the bundle containing 30 residues (Aα166 to Aα195) that begins at the lateral end of the coiled-coil where the Aα chain goes in the reverse direction for about ¼ of the length of the coiled-coil, at which point the rest of the C-terminal part of the Aα chain becomes outstretched (Spraggon et al. 1997). The coiled-coils can bend around a hinge point located in the γ chain near the residue γAsn52 shown in Fig. 15.1 (Marsh et al. 2013). The hinge points of the coiled-coils are highly susceptible to plasmin cleavage, resulting in formation of fibrinogen's fragments. The hinge also determines the flexibility of fibrinogen molecules observed at interfaces (Köhler et al. 2015) or in solution (Zuev et al. 2017), and has a biomechanical role discussed below.

Carbohydrate Moiety

Each fibrinogen molecule has four N-glycans connected to the symmetrical BβAsn364 residues (located in the β-nodules) and γAsn52 (located in the coiled-coils), the latter carbohydrates not resolved crystallographically (Fig. 15.1). The presence and structural variations of the oligosaccharides affect fibrinogen's solubility, life span in the blood stream, kinetics of fibrin polymerization, and clot structure.

Conversion of Fibrinogen to Fibrin

Formation of Monomeric Fibrin and A:a Knob-Hole Interactions

Fibrin formation is initiated by the thrombin-catalyzed release of two 16-residue-long peptides (FpA) from the N-termini of the Aα chains of fibrinogen to produce desA-fibrin monomer. The release of two 14-residue-long peptides (FpB) from the N-termini of the Bβ chains occurs at a much slower rate and is not necessary for fibrin polymerization. The critical importance of FpA release is supported by dysfibrinogenemias or genetic fibrinogen variants with impaired cleavage of FpA that precludes the enzymatic conversion of fibrinogen to fibrin (Galanakis et al. 1989).

The enzymatic cleavage of FpA in fibrinogen results in exposure of a new glycine(Gly)-proline(Pro)-arginine(Arg) motif. This tripeptide at the N-terminus of fibrin's α chain is called knob 'A' to highlight that it fits into a complementary hole 'a' located in another monomeric fibrin, hence their binding is named the A:a knob-hole interaction (Fig. 15.2). The first evidence for the high affinity and structural compliance of knobs 'A' and holes 'a' was provided by X-ray crystallography of fibrinogen fragment D co-crystallized with the free peptide Gly-Pro-Arg-Pro that mimics the sequence of knob 'A'. Because fragment D corresponds to the lateral D region, the binding pocket for the synthetic knob 'A' was identified as the hole 'a' located in the γ-nodule (Everse et al. 1998b; Kostelansky et al. 2002). If holes 'a' are blocked by synthetic Gly-Pro-Arg-Pro (Everse et al. 1998b) or compromised by a point mutation of the key residue γAsp364 (Okumura et al. 1997), fibrin polymerization is impossible. Overall, these and other data indicate that A:a knob-hole interactions, mainly electrostatic by nature, are the driving force of fibrin polymerization and blood clot formation.

The natural A:a knob-hole interactions between full-length fibrin(ogen) molecules have been reproduced and measured at the single-molecule level and these A:a complexes turned out to be strong and mechanically stable (Litvinov et al. 2005). Surprisingly, in response to increasing pulling force an unusual strengthening of A:a knob-hole bonds followed by their weakening was revealed, suggesting the existence of A:a complex in two distinct conformational states (Litvinov et al. 2018). The amino acid residues in holes 'a' binding to Gly-Pro-Arg-Pro are located within a segment γ290-379 and include γAsp364, γArg375, γHis340, γGln329, and γLys338 (Everse et al. 1998b). However, the interaction between fibrin monomers involves residues other than those comprising the "hot spots" in holes 'a' and knobs 'A'. In particular, at the D:E:D interface residues αVal20-Lys29 (adjacent to the knob 'A' sequence αGly17-Arg19) interact with residues γAsp298-Phe304, γAsn319-Asn325, and γGln329 in the γ-nodule area next to hole 'a'. There are also binary contacts between residues αLeu54, αGlu57, αPhe62, αArg65, γGln33, γAsp37, and γGln49 in the coiled coil and γGln329-Asp330 in the γ-nodule (Zhmurov et al. 2016). In other words, in the real monomer-monomer interactions there is an intermolecular interface beyond knobs 'A' and holes 'a' that can be considered more generally as *binding sites* '*A*' and '*a*'.

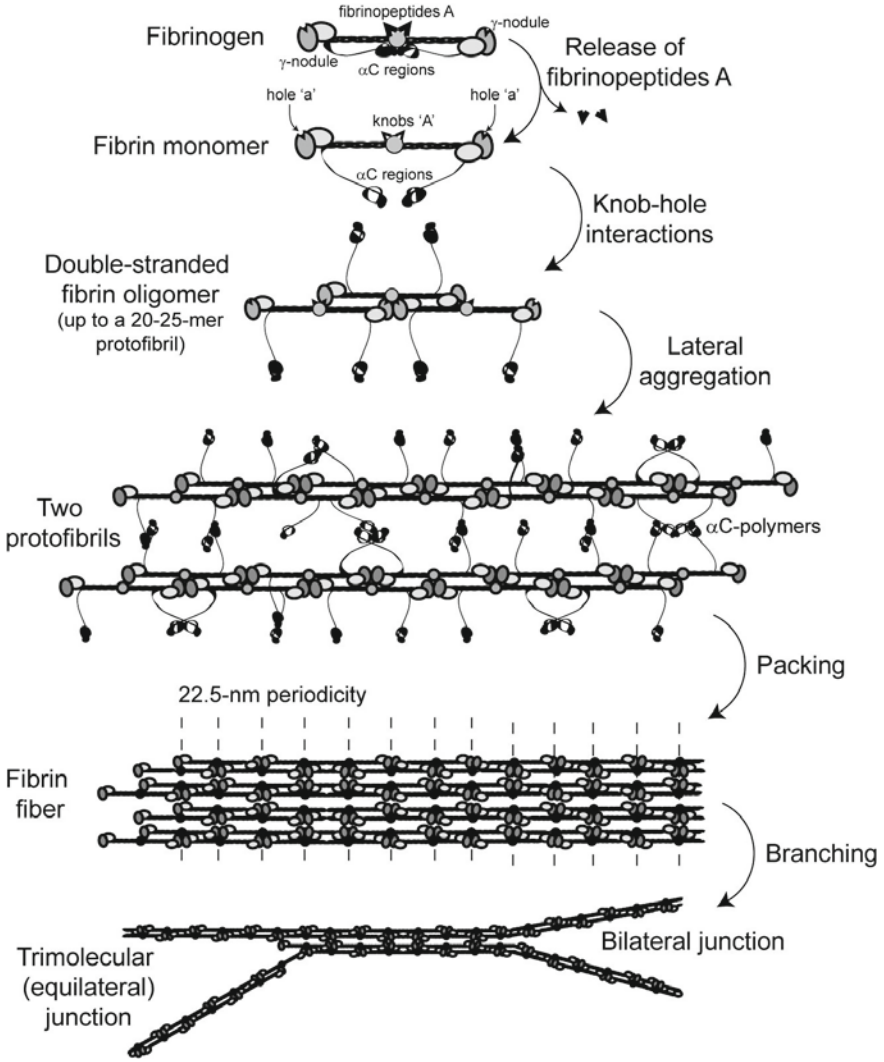


Fig. 15.2 Schematic diagram of fibrin polymerization. Fibrinopeptides in the central nodule cover knobs that are complementary to holes that are always exposed at the ends of the protein. When the fibrinopeptides are removed by thrombin, knob-hole interactions occur, giving rise to oligomers (a trimer is shown), which elongate to produce the two-stranded protofibrils made up of half-staggered molecules. The protofibrils aggregate laterally to make fibers, a process enhanced by interactions of the αC regions and formation of the αC-polymers. The fiber has a 22.5 nm periodicity as a result of half-staggering of 45 nm molecules. At the bottom of the diagram, branch points have been initiated by the divergence of two protofibrils (right) and splitting of each strand of a single protofibril (left) (Weisel and Litvinov 2017)—with permission from Springer Nature

Formation of Fibrin Protofibrils

As the first non-enzymatic step of fibrin polymerization, highly reactive desA-fibrin monomers (i.e. molecules with both FpA cleaved off) self-assemble in a half-staggered manner so that two knobs 'A' located in the central E region of one elongated molecule fit into two holes 'a' located in the D regions of two rod-like monomers connected end-to-end (Erickson and Fowler 1983). This primary polymerization step leads to formation of a two-stranded trimer, in which the lateral D regions of two linearly arranged molecules form the D:D interface and both get tethered via knobs 'A' to the E region of the third monomer in the other strand. Therefore, a triple D-E-D complex comprises the junction between three monomers in two strands held together mainly by the A:a knob-hole bonds (Fig. 15.3) (Everse et al. 1998a). The D:D interface is formed by residues γ275-309 (Everse et al. 1998b) of which residues γ275, γ308, and γ309 are essential for addition of new monomers to each strand to form a longer linear fibrin oligomer (Marchi et al. 2006; Bowley et al. 2009). The D:D interfacial contacts form a hinge, which is localized to residues γAla279-Tyr280 in one D region and residues γAsn308-Gly309 in the other. This hinge allows bending at the D-D interface by ~ 60° or more (Zhmurov et al. 2016). Importantly, point mutations in the residues γAla279, γTyr280, γAsn308, and γGly309, which stabilize the D-D interactions, are implicated in a number of dysfibrinogenemias with impaired fibrin formation (Hanss and Biot 2001). The oligomeric strands grow longitudinally until they reach a length of a double-stranded protofibril (Fig. 15.2), a soluble intermediate product of fibrin polymerization, comprising a straight twisted filament about 0.5–0.6 μm in length made of about 20–25 monomers (Erickson and Fowler 1983; Chernysh et al. 2011). The persistence length of a protofibril is about 320 nm with a high propensity for bending and kinking (Zhmurov et al. 2018). What makes protofibrils critically important in fibrin polymerization is that they reach a threshold length to undergo lateral aggregation and thus form thicker fibrin fibers. An unusual mechanism for the early steps of fibrin polymerization has been proposed that implies formation of single-bonded 'Y-ladder' polymers that grow linearly before undergoing a transition to the double-stranded protofibrils (Rocco et al. 2014).

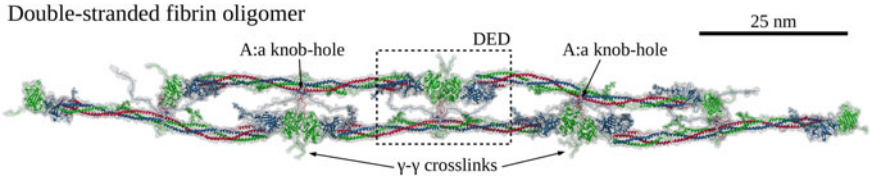


Fig. 15.3 Structure of a double-stranded fibrin oligomer composed of five monomers stabilized mainly by the A:a knob-hole interactions that form the D-E-D complex Modified from (Zhmurov et al. 2018)—published with permission from Elsevier

Formation of Fibrin Fibers

Fibrin fibers are formed as a result of two simultaneous and competing processes: further linear elongation of protofibrils and their lateral (or radial) aggregation that determine the variability of the fiber mass/length ratio. Depending on the conditions of fibrin formation, the number of protofibrils per fiber cross-section can vary from about two to several thousand (Missori et al. 2010; Yeromonahos et al. 2010). During regular side-to-side bundling, protofibrils form fibers with the axial periodicity of ~ 22.5 nm (½ of the length of a fibrin monomer), originating from the precise half-staggered molecular packing (Fig. 15.2). This localized protein densification at the transversely aligned globular D-E-D complexes is visualized as periodic cross-striation of fibrin fibers with transmission electron microscopy and atomic force microscopy (Fig. 15.4) (Weisel 1986; Yermolenko et al. 2011; Zhmurov et al. 2018). The molecular arrangement within a fibrin fiber is highly regular in the longitudinal direction, while lateral packing is less ordered (Torbet et al. 1981; Weisel et al. 1983; Weisel 1986; Caracciolo et al. 2003; Yeromonahos et al. 2010; Jansen et al. 2020).

The estimated spatial density of protofibrils varies depending on conditions of fibrin formation and methods applied (Yeromonahos et al. 2010). For example, increased thrombin concentration has been demonstrated to lead to decreased fiber radius and a markedly lower average protofibril content per fiber. This resulted in the formation of less compact, less dense fibrin structures, with distances between protofibrils within the fiber being larger at higher thrombin concentrations (Domingues et al. 2016). In addition, a common fibrinogen splice variant, named γ' fibrinogen, which results from an alternative polyadenylation signal in intron 9, resulting in a γ-chain with a 16-amino-acid C-terminal extension, has also been shown to alter protofibril packing (Domingues et al. 2016). In plasma,

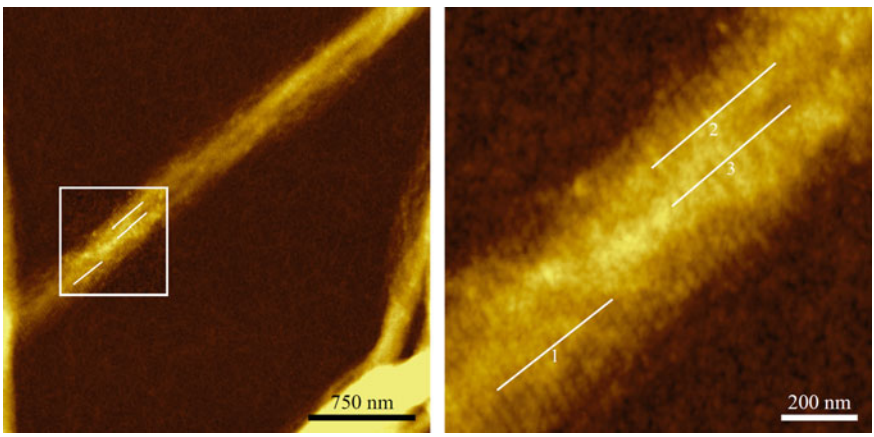


Fig. 15.4 Cross-striation of a fibrin fiber with a 22.5 nm periodicity visualized using atomic force microscopy (Yermolenko et al. 2011)—with permission from ACS Publications

γ' fibrinogen mainly circulates as a heterodimer with the more common γA-chain (γA/γ') (Wolfenstein-Todel and Mosesson 1981), with an average plasma concentration of 8–15% (Mosesson et al. 1972; Chung and Davie 1984). Domingues et al. (2016) demonstrated a reduction in protofibril packing for fibers produced using γA/γ' compared to γA/γA fibrinogen, across a range of thrombin concentrations. The γ' chain itself seems to influence polymerization, with specifically lateral aggregation of protofibrils being slowed down and the formation of non-uniform clots consisting of thin fibers (Gersh et al. 2009b; Allan et al. 2012). Protofibril packing is decreased in fibers containing γ' fibrinogen, with protofibrils being packed loosely and fibers unraveling into loose fibrils with diameters of less than 10 nm, consistent with the diameter of individual protofibrils (Domingues et al. 2016). Fibers with reduced protofibril packing either due to increased thrombin concentration or the presence of γ' fibrinogen are less stiff than more dense fibers, leading to a weakened clot architecture (Domingues et al. 2016) (Fig. 15.5). The importance and structural basis of clot mechanical properties are discussed in the next section.

Based on indirect evidence, it has furthermore been proposed that the lateral packing of fibrin fibers is non-uniform and that they have a dense, relatively compact core and sparse, loosely arranged protofibrils at the periphery (Li et al. 2017). Irrespective of the extent of lateral order and uniformity of protofibril packing, the protein density within fibrin fibers was estimated to be around 20–30% (Yeromonahos et al. 2010), indicating that fibers have large inter-protofibril spaces filled with liquid. Additionally, using small-angle X-ray scattering, the spatial density of protofibrils has been recently estimated to have a typical repeat distance of ~ 13 nm, independent of the fiber thickness or the number of protofibrils per fiber cross-section (Jansen et al. 2020).

Each protofibril is twisted; therefore, fibrin fibers formed during lateral aggregation of the axially oriented protofibrils are also twisted (Medved et al. 1990) with a 20-60 nm helical radius and ~ 2000 nm helical pitch (Weisel et al. 1978). New protofibrils that are added to the surface of a fiber must be stretched as their

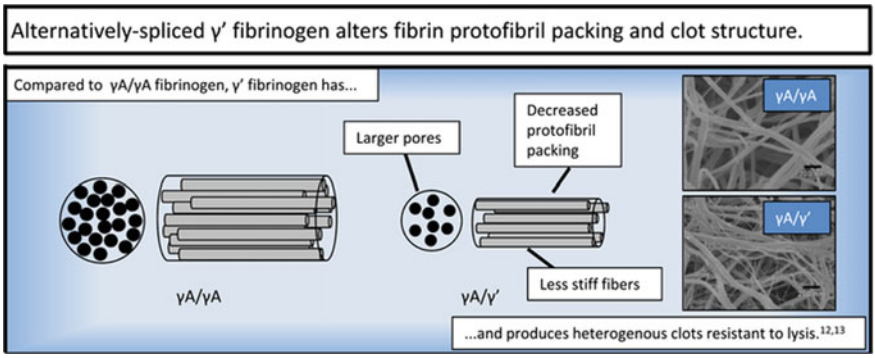


Fig. 15.5 Alternatively spliced γ' fibrin(ogen) alters protofibril packing and clot structure (Pieters and Wolberg 2019)

path length increases with fiber thickness. The twisting deformation of protofibrils provides a thermodynamic mechanism to limit the thickness of fibrin fibers, as they stop growing laterally when the energy necessary to stretch an added protofibril is equal to or exceeds the energy of attachment.

The particular binding sites mediating the lateral aggregation of protofibrils are unknown, but a number of structures have been indirectly shown or assumed to mediate the associations of protofibrils, namely the αC regions, B:b knob-hole complexes, the C-terminal portions of the γ chains and two adjacent β-nodules (Yang et al. 2000), the coiled-coil connectors (Okumura et al. 2006), and carbohydrate moieties (Langer et al. 1988). There are conceivable structural models for the lateral aggregation of protofibrils based on the interactions between two β-nodules of adjoining protofibrils via residues β330-375 (Yang et al. 2000) and on the involvement of the N-terminal part of the β chain, namely residues βAla68 (Mullin et al. 2000) and βGly15, the end residue of knob 'B', irrespective of whether FpB is cleaved or not (Hirota-Kawadobora et al. 2003). Based on the structural models of the compact part of the fibrinogen α chain and the N- termini of the β chain, their involvement in lateral aggregation of protofibrils was suggested by formation of intricate lattices together with the γ-chain (Klykov et al. 2020). In addition, detailed structural analysis of dysfibrinogenemias with impaired clot formation has revealed a crucial role for the βArg166 residue in lateral aggregation (Klykov et al. 2020).

Additional Molecular Interactions During Fibrin Polymerization

Besides the essential A:a and D:D associations, there are a number of intermolecular interactions that are not necessary for fibrin polymerization but they can modify the ultimate structure and properties of fibrin, such as B:b knob-hole interactions, homomeric oligomerization of the αC regions, and covalent crosslinking catalyzed by the plasma transglutaminase, factor XIIIa.

The thrombin-catalyzed release of FpB from the N-termini of β-chains exposes a new Gly-His-Arg-Pro motif, which is called knob 'B' that fits into hole 'b' located in the β-nodule, a part of the globular D region. The B:b interactions were reproduced and measured in vitro (Litvinov et al. 2006) but the physiological role of the B:b interactions remains uncertain. In the absence of knobs 'B', when thrombin cleaves off FpA from fibrinogen while release of FpB is prevented, fibrin has thinner fibers than those formed after cleavage of both fibrinopeptides, suggesting that B:b interactions contribute to lateral aggregation of protofibrils (Blombäck et al. 1978). The B:b interactions in fibrin are real because some homodimeric dysfibrinogenemias (fibrinogen variants Metz and Frankfurt XIII) from which only FpB can be cleaved (no FpA release), produce fibrin clots at low temperature exclusively via B:b bonding (Galanakis et al. 1993; Galanakis 1993). Accordingly, replacement of the γAsp364 residue in fibrinogen, making holes 'a' dysfunctional and preventing formation of

A:a complexes, still allows fibrin clots to form slowly under the action of thrombin (Okumura et al. 2007), again confirming existance of B:b interactions in the absence of A:a bonding. Notably, holes 'b' can interact with knobs 'B' both within and between protofibrils, but, due to steric constraints, the intra-protofibril B:b contacts are less probable than the inter-protofibril B:b contacts, which explains why the B:b interactions are not necessary for initiation of fibrin polymerization and why they modulate the lateral aggregation of protofibrils (Zhmurov et al. 2018). In addition to their potential role in lateral aggregation of protofibrils, B:b interactions change the susceptibility of fibrin to enzymatic proteolysis (Doolittle and Pandi 2006). The functional importance of B:b knob-hole interactions was confirmed indirectly by formation of fibrin with distinct structure and properties in the presence of a knob 'B' mimetic peptide (Gly-His-Arg-Pro) conjugated with polyethyleneglycol (Brown et al. 2015).

The extended and flexible αC regions of fibrin(ogen) molecules can self-interact within and between protofibrils, resulting in formation of αC polymers (Fig. 15.2) (Tsurupa et al. 2011; Protopopova et al. 2017). Similar to B:b knob-hole interactions, the αC regions promote lateral aggregation while being not essential (Weisel and Medved 2001; Tsurupa et al. 2011). Fibrin formed from des-αC-fibrinogen or fibrinogen with short αC regions has thinner fibers with more branch points than fibrin formed from the full-length fibrinogen, suggesting impaired lateral aggregation of protofibrils in the absence of long enough αC regions (Collet et al. 2005; Ping et al. 2011). However, in the absence of functional knobs 'A' and 'B' and therefore without A:a and B:b knob-hole interactions the αC-regions' interactions appear too weak and unstable to mediate association of fibrin(ogen) molecules (Duval et al. 2020).

During and after polymerization, fibrin is covalently crosslinked by factor XIIIa, an active transglutaminase formed from inactive factor XIII by thrombin in the presence of Ca^{2+}. The first sites to be crosslinked are located in the unstructured C-terminal tails of γ chains approximated due to D-D contacts between two linearly adjacent fibrin monomers within the same strand (Fig. 15.6). Factor XIIIa catalyzes formation of an intermolecular ε-(γ-glutamyl)-lysyl isopeptide bond between the γLys406 of one γ chain and γGln398/399 of another γ chain. Factor XIIIa forms the same type of bonds at a slower rate between lysine and glutamine residues in the αC regions associated into αC polymers within protofibrils and in the inter-protofibril space (Matsuka et al. 1996). The fibrin α and γ chains also can be crosslinked, resulting in formation of α-γ-heterodimers (Standeven et al. 2007). The covalent stabilization makes fibrin polymerization irreversible and the crosslinked fibrin clot becomes mechanically stiff and more resistant to fibrinolysis.

Formation of a Fibrin Network

As fibrin fibers thicken due to lateral aggregation of protofibrils and grow in length due to longitudinal addition of new monomers or oligomers, they also branch, which

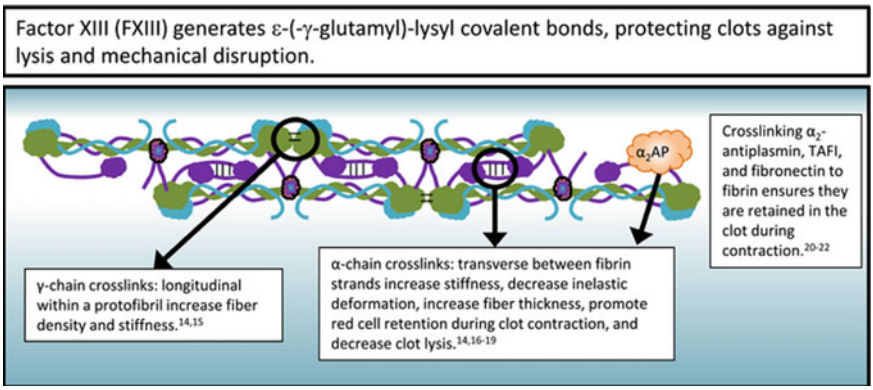


Fig. 15.6 Formation of isopeptide bond catalyzed by factor XIIIa. The chemical reaction catalyzed by Factor XIIIa, yielding insoluble fibrin crosslinked by ε-(γ-glutamyl)-lysine bonds between the γ chains and the αC regions (Pieters and Wolberg 2019)

is necessary to yield a three-dimensional network (Fig. 15.2). There are at least two types of branching points formed by distinct mechanisms that may both contribute to fibrin network structure. The one called a "bilateral junction" appears when two protofibrils bundle to form a four-stranded filament and then diverge again into two separate protofibrils (Mosesson et al. 1993). The second type of branching is called a "trimolecular junction" or "equilateral junction" because it arises when a fibrin monomer binds at the end of a protofibril via only one γ-nodule, such that this partial bimolecular complex can initiate formation of a new two-stranded protofibril (Fogelson and Keener 2010). Irrespective of the mechanism, most branch points in fibrin networks consist of three fibers of about the same diameter joined together (Ryan et al. 1999). Generally, there is an inverse relation between the number of branch points in a clot and the fiber diameters (Ryan et al. 1999). In other words, in fibrin clots with thick fibers there are few branch points, while clots made up of thin fibers have many branch points. Because branching occurs during elongation of protofibrils, while fiber diameter is determined by the lateral aggregation, it can be inferred that branching and lateral aggregation compete. Alternative to the commonly accepted three-dimensional fibrin network structure, one not universally recognized, but experiment-based model of fibrin polymerization includes formation of ultrathin fibrin sheets (O'Brien et al. 2008).

As a result of combined elongation, branching, and thickening of fibrin fibers, a space-filling network is formed that macroscopically represents an insoluble hydrogel. The time point at which the gel is formed is named the clotting time of blood or plasma and used in laboratory assays as a quantitative parameter to reveal coagulation disorders, often associated with prolonged or accelerated clotting time. The gelation point corresponds to the state of clotting when only about 15–20% of soluble fibrinogen has been incorporated into the insoluble gel, implying that the fibrin network is not yet established at the time of gelation, with new fibers and branch points still being formed after the gelation point (Chernysh and Weisel 2008). The

structure of ultimate fibrin clots is quantified using scanning electron microscopy, confocal microcopy, and optical spectroscopy by the mass/length ratio, diameters, and spatial density of fibers, number of branch points and distances between them (fiber length), as well as the size of the pores. All of these parameters are strongly affected by variations in the kinetics of polymerization determined by thrombin and factor XIIIa activity, the quality and quantity of fibrinogen, the presence of blood cells as well as numerous environmental factors. It is noteworthy that in vitro fibrin networks formed in static conditions at physiological fibrinogen concentrations are very porous since the mass fraction of fibrin is about 0.3%, corresponding to the normal fibrinogen level in plasma. By contrast, when a clot in formed in dynamic conditions of blood flow, the composition and structure of thrombi and thrombotic emboli, including the mass fraction and structural diversity of fibrin, are dramatically different (Chernysh et al. 2020).

Mechanical Properties of Clots

Significance of Clot Mechanical Properties

Although the mechanical properties of clots are often ignored, they are among the most important functional characteristics (Weisel 2004). In hemostasis, clots must form a seal to stem bleeding, including the force of arterial blood flow. In thrombosis, the mechanical properties of clots determine the fate of thrombi. The pressure of flowing blood can dislodge or deform a thrombus or cause it to rupture, a process called embolization, which is responsible for pulmonary embolism and some forms of ischemic stroke. Furthermore, there is now strong evidence for a correlation between clot stiffness and many (pro) thrombotic conditions (Collet et al. 2006).

Basis of Clot Mechanical Properties

Fibrin is the scaffold of clots and thrombi that is largely responsible for their mechanical properties. It forms a network that is space-spanning to support forces applied to a clot. While some proteins, like collagen or actin form fibers or filaments, they do not form a gel without the binding of other proteins. Other proteins, such as elastin and gelatin, make networks without forming fibers. Fibrin fibers form networks because the fibers branch, such that fibers diverge from each other, usually with three fibers joining at each branch point. In fact, branching is such a dominant feature of fibrin clots that fiber ends are rarely seen in clots made from normal fibrinogen (Weisel 2004).

Clots that include platelets, such as clots made from recalcified platelet-rich plasma, have a different architecture to clots from platelet-poor plasma. Since most

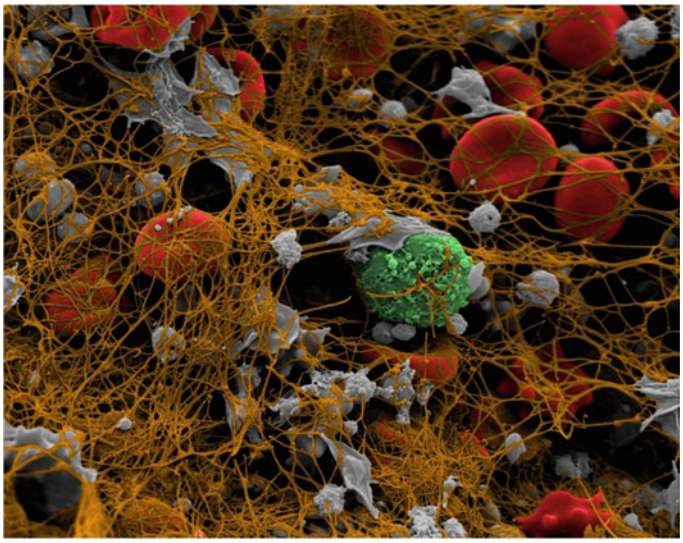


Fig. 15.7 Colorized scanning electron micrograph of a human coronary thrombus, showing fibrin (orange), platelets (grey), erythrocytes (red), and a leukocyte (green) (Weisel lab images)

of the thrombin is generated on the surface of the platelets, that is where fibrinogen is converted to fibrin, and hence many fibers originate from platelet aggregates, radiating like the spines of a starfish (Collet et al. 2002). Clots made from whole blood include red blood cells, which occupy a considerable volume of the clot and increase the pore size of the fibrin mesh (Gersh et al. 2009a). Clots formed in vivo can have a quite different architecture, because of the spatial initiation of clotting and because of blood flow (Fig. 15.7). Thrombi are also quite different than either in vitro clots or in vivo clots, since they are usually formed intravascularly as a result of pathological processes. However, although all these conditions have distinctive effects on clot or thrombus properties, fibrin is the structural and mechanical basis of all of them (Liang et al. 2017).

Viscoelastic Properties of Fibrin

Fibrin is called a viscoelastic polymer because it has both elastic properties, which means that strain or deformation is proportional to stress or force per area, and viscous properties, which means that stress is proportional to the rate of strain. Elastic properties, stiffness or elastic modulus or storage modulus, characterize the reversible deformation by force, while the viscous properties, plasticity or loss modulus, characterize the irreversible or inelastic properties (Weisel 2004).

Commonly, rheometers are used to measure the viscoelastic properties of fibrin. Usually shear rheometers, which impose a shear stress or shear strain on a clot, are

employed, but some rheometers can also be used for compression, or others for tensile testing, or stretching. In addition, there are specialized devices in the clotting field that are not rheometers but measure clot stiffness over time and yield parameters that are used by clinicians to evaluate the formation of clots from plasma or whole blood (Luddington 2005; Ganter and Hofer 2008; Bolliger et al. 2012). These results of thromboelastography or thromboelastometry, which are more global ways to measure clot firmness, are used to diagnose patients with bleeding or thrombotic disorders.

For fibrin clots, the elastic component is generally about an order of magnitude greater than the viscous component, although the viscous component increases rapidly with frequency. Stiffness is greater for clots made from the plasma of patients with thrombotic conditions (Collet et al. 2006), while it is less for patients with hemophilia A (Leong et al. 2017).

Non-linear Elasticity and High Extensibility of Fibrin

Stress-strain curves, with applied stress plotted against the strain or induced deformation, obtained using a rheometer can be used for characterizing the elasticity of clots. The slope of the stress-strain curve is the stiffness of the clot. Fibrin exhibits strain stiffening or strain hardening, which means that the slope of the curve becomes greater with increasing strain (Storm et al. 2005; Janmey et al. 1983).

Since fibrin is highly extensible, clots will tend to stretch rather than break. Fibrin in a plasma clot that is crosslinked with Factor XIIIa can be stretched about threefold before rupture (Brown et al. 2009). Fracture resistance of fibrin has not been much studied, but it depends strongly on the red blood cell content, which may be significant in vivo for embolization (Riha et al. 1999). Remarkably, the strain stiffening and extensibility of fibrin clots is also observed in stretching of individual fibrin fibers (Liu et al. 2006; Guthold et al. 2007; Liu et al. 2010). The elastic modulus of individual fibrin fibers is strongly dependent on the fiber diameter, indicating that the packing of the protofibrils is different in thick and thin fibers, as discussed above (Li et al. 2016). The strain stiffening in individual fibers allows the strain load to be distributed more evenly over the network of fibers, from the highly strained and hence stiffer fibers to less strained fibers (Hudson et al. 2010).

The non-linear elasticity of clots has also been observed in compression. Initially, fibrin clots decrease in stiffness with compression, as fibers perpendicular to the direction of compression bend and buckle (Kim et al. 2014). With further increasing compression, there is a striking non-linear increase in clot stiffness, as fibers parallel to the direction of compression encounter each other with an increase in network density from criss-crossing fibers.

The physiological significance of fibrin stiffening likely arises because clots will deform in response to forces, but then be protected from damage when they become stiffer at higher forces. Furthermore, mechanical stress and the resulting deformation of clots make them more resistant to fibrinolysis, which may be one reason for lesser efficacy of thrombolytic treatments of thrombi in some conditions. Moreover, the

combination of histone-DNA complexes mimicking neutrophil extracellular traps, and fibrin makes those clots stiffer and more resistant to fibrinolysis (Longstaff et al. 2013).

Multiscale Fibrin Mechanics

The mechanical properties of clots can only be understood by consideration of their structure at a wide range of spatial scales, including macroscopic, fiber network, individual fiber, and molecular levels (Fig. 15.8). At the macroscopic scale, fibrin's strain stiffening and high extensibility have already been discussed. In addition, clots display a large decrease in volume with both stretching and compression, which is due to water expulsion and network densification (Brown et al. 2009). One reason for this phenomenon may be negative normal stress, because fibers buckle more easily than they stretch (Kang et al. 2009). However, a more significant explanation is likely due to a molecular structural transition that will be discussed further below.

At the network level, plasma clots have a strikingly porous structure with a mass fraction of protein of only about 0.25%. In addition, the structure of these clots is highly dependent on the conditions of polymerization. High thrombin concentrations make clots with thin fibers and many branch points that are stiffer than clots at lower thrombin levels, while at low ionic strength or high calcium ion concentration or

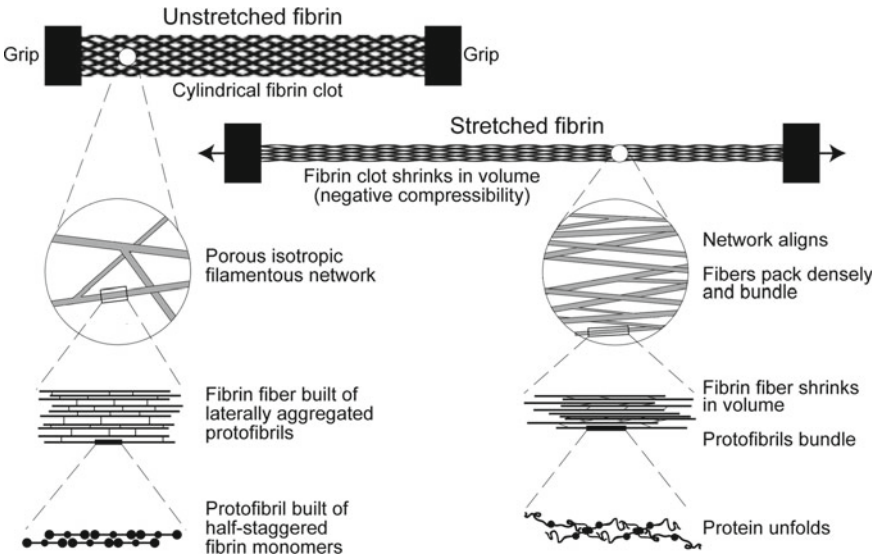


Fig. 15.8 The diagram shows the structural changes that occur upon stretching the cylindrical fibrin clot at different spatial scales, including macroscopic volume shrinkage, network rearrangement, and protein unfolding (Purohit et al. 2011)—with permission from Elsevier

high pH, clots have thicker fibers and fewer branch points and are softer (Weisel and Litvinov 2013; Weisel and Litvinov 2017). All of these clots are initially made up of a branching isotropic network, with no preferred orientation of fibers. With stretching, the fibers align along the direction of strain. In addition, the fibers become thinner and self-associate. With compression, the fibers become oriented in the direction perpendicular to the direction of compression and become shorter because of criss-crossing (Kim et al. 2014). As a result, the shear modulus or stiffness of the clot in response to shear strain, increases dramatically, since the fibers are now oriented and more highly branched.

At the fiber level during stretching, fibers are increasingly under large strains after they become oriented. Fibers under compression bend and buckle in the direction of applied stress. As more fibers buckle, there are fewer fibers to sustain loads, so the stiffness decreases.

In recent years, there have been a plethora of studies of clot mechanics at the molecular level to such an extent that a description of these results deserves its own section, which follows.

Molecular Structural Basis of Fibrin Mechanics

With stretching of clots, after the initial reorientation and alignment of fibers along the direction of applied force, there must be changes in the molecular structure of fibrin. There are three major candidates for such molecular changes: unfolding of the γC regions, unfolding of the α-helical coiled coil, stretching of the mostly unstructured connector of the αC region. There is still disagreement on the roles that each of these structural changes play, but it now seems likely that all occur to at least some extent. From the crystal structure of fibrinogen and taking into account the disulfide bonds that restrict the extent of stretching, it can be predicted that hypothetically the full unfolding of the molecule would result in a 4.7-fold elongation (Zhmurov et al. 2011).

Small angle X-ray scattering (SAXS) patterns of fibrin are characterized by a periodicity of 22.5 nm, because of the half-staggering of 45 nm molecules. It was expected that stretching of clots would result in an increase of this periodicity. Although some early studies appeared to show such an increase, it now seems clear that there is little if any increase in periodicity (Brown et al. 2009). However, there is a broadening of the 22.5 nm peak, indicating that there is an increase in disorder of the packing. These results are consistent with stochastic unfolding of an increasing number of molecules with strain. This behavior was replicated with a two-state model in which some molecules randomly extend and unfold while others remain folded (Brown et al. 2009).

Single molecule force spectroscopy, particularly using atomic force microscopy, has been used to study protein unfolding and characterize unfolding transitions through analysis of force-distance curves. However, it is nearly impossible to interpret the results from such experiments with fibrinogen molecules or naturally

occurring fibrin polymers, because these structures are quite complex and many different unfolding events are possible. As a partial solution, tandems or single-stranded fibrinogen oligomers were prepared by cross-linking of fibrinogen end-to-end with Factor XIIIa (Zhmurov et al. 2011). An additional part of the solution to understanding the experimental results has been to combine them with Molecular Dynamics simulations and modeling of the unfolding transitions from the known X-ray crystallographic structures. With this dual approach, it was discovered that there is unfolding of the C-terminal γC nodules, accompanied by reversible extension and refolding of coiled-coil connectors, which act as springs to take up the slack from unfolding of globular regions (Zhmurov et al. 2011).

Since fibrinogen is a fibrous protein with two relatively long axially aligned α-helical coiled-coils, they are an obvious candidate for unfolding, especially since it has been demonstrated that other fibrous proteins, such as keratin can unfold with applied force. The α-helical coiled-coils in fibrin undergo a transition from α-helix to β-sheet with either stretching or compression (Fig. 15.9) (Litvinov et al. 2012).

There is considerable evidence for the contribution of the αC regions to clot mechanical properties. Factor XIIIa crosslinking of the αC regions makes clots much stiffer and decreases their plasticity (Collet et al. 2005). Studies of fibrinogens from different species with different lengths indicates that the contribution of this part of the molecule is important (Falvo et al. 2008). In addition, there is evidence that the fast elastic recoil of fibrin fibers may arise from the αC regions (Hudson et al. 2013). Study of changes in the SAXS pattern of fibrin with shear strain also show little change in periodicity. From analysis of the forces applied to fibrin molecules and the measured unfolding response of fibrin to forces, it has been proposed that in these experiments there was no unfolding of domains, but instead stretching of the unstructured part of the αC region (Vos et al. 2020).

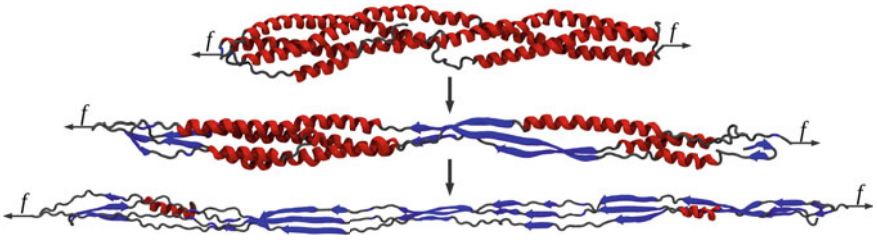


Fig. 15.9 Force-induced mechanical transition of α-helices (red) into β-sheets (blue) in the fibrin(ogen) triple-helical coiled-coils (Weisel and Litvinov 2017)—with permission from Springer Nature

Modeling Fibrin Mechanical Properties

Models of the mechanical properties of filamentous networks are one way to take experimental data and use it to understand and predict fibrin mechanics. One such model uses a system of connected fibers with known force-extension curves. Although the fibrin network is not modeled explicitly in terms of its known structure, this model captures the stress-strain relationship of fibrin clots, using the measured force-extension curves for individual fibers (Brown et al. 2009). In this model, the fibrin molecules are considered to be a two-state system that can either be folded, where it's modeled as a linear spring, or unfolded, where it's a worm-like chain. This model fits the experimental data very well, if the observed decrease in volume is incorporated into the model.

Another model assumes that fibrin fibers are semiflexible and are subject to thermal fluctuations that are reduced in response to stretching (Storm et al. 2005). This model can also be used to predict clot mechanical properties, starting with known force-extension curves for individual fibers. Several models utilize the concept of wormlike semiflexible polymers, some of which do not include the idea of fibrin unfolding, which may depend on the type and degree of deformation (Piechocka et al. 2010). Other models only described the mechanical properties of individual fibers and do not attempt to model the clot network. Models of compression and shear often differ from those for extension. With both compression and tension, criss-crossing of fibers and strong interactions between them are an important part of the models (Kim et al. 2014; Britton et al. 2019). Some models for compression of fibrin clots use the theory of foams or cellular solids (Kim et al. 2016). Other models have considered non-affine deformations that occur in certain types of clots, depending on the conditions of formation (Kang et al. 2009). Finally, an additional level of complexity is the consideration of cyclic loading, which is more complex because irreversible changes occur (Liang et al. 2017).

Clot Properties and Disease

The primary function of fibrin clot formation, is to prevent unwanted blood loss by forming, together with platelet adhesion and aggregation and other blood cells, a stable plug at the site of vascular injury. The structural and mechanical fibrin clot properties described above are critical for successful hemostasis. Alteration in these properties can consequently result in pathological vascular coagulation, such as thrombosis or alternatively, bleeding. The consistent findings of altered clot properties in both venous and arterial thrombotic conditions provide a plausible mechanistic link between clot structure and thrombosis, although causality remains to be confirmed. In keeping with the Mendelian randomization approach, there is some evidence linking a number of fibrinogen variants that alter fibrin clot properties

to cardiovascular diseases. In addition, recent reports providing evidence that altered clot properties predict the development of cardiovascular complications and even malignancy, may provide support for its potential causal role in disease.

Fibrin Clot Properties and Pathological Intravascular Coagulation

Clots with densely-packed, thin fibers, with a stiffer network arrangement, reduced permeability (smaller pore sizes) and enhanced resistance to fibrinolysis have consistently been found in a variety of cardiovascular and other diseases (Fig. 15.10). Examples are:

- Arterial thrombosis: ischemic stroke (Undas et al. 2010; Undas and Ariens 2011; Bridge et al. 2014), coronary artery disease (Mills et al. 2002; Collet et al. 2006; Undas and Ariens 2011; Bridge et al. 2014; Undas 2016), peripheral arterial disease (Undas and Ariens 2011; Litvinov and Weisel 2016; Undas 2016), acute coronary syndrome (Undas et al. 2008; Undas 2016), no-reflow phenomena after acute MI (Undas 2016), and in-stent thrombosis (Undas 2016; Kattula et al. 2017).
- Venous thrombosis and thromboembolism: deep vein thrombosis, pulmonary embolism (Bridge et al. 2014), and cerebral venous sinus thrombosis (Undas 2016). Differences in clot properties between deep vein thrombosis patients not developing thromboembolism and those who do, suggest a potential role for clot properties in embolism risk (Bridge et al. 2014).
- Chronic inflammatory diseases: inflammatory bowel disease (Undas 2016), antiphospholipid syndrome (Undas 2016), rheumatoid arthritis (Undas and Ariens 2011; Ariens 2013), chronic obstructive pulmonary disease (Undas and Ariens 2011; Ariens 2013), and systemic lupus erythematosus (Litvinov et al. 2019). Fibrin plays an important role in innate immunity and can contribute to inflammation by recruiting inflammatory cells and enhancing leukocyte reactivity by binding to the Mac-1 integrin (Trezzini et al. 1988; Altieri et al. 1990; Flick et al. 2004).
- Other: chronic heart failure with sinus rhythm (Palka et al. 2010), atrial fibrillation (Undas 2016), arterial hypertension (Undas 2016), aortic aneurysm (Ariens 2013; Bridge et al. 2014), disseminated intravascular coagulation (Litvinov and Weisel 2016), diabetes mellitus (Bridge et al. 2014; Undas 2016), end stage renal disease (Ariens 2013; Undas 2016), cancer (Undas 2016), and liver diseases such as cirrhosis (Hugenholtz et al. 2016), in part due to intrinsic changes in the fibrinogen molecule such as post-translational modifications (Lisman and Ariens 2016; De Vries et al. 2020).
- Atherosclerosis – fibrin deposits in plaque contribute to plaque growth and (in) stability (Lepedda et al. 2009; Borissoff et al. 2011; Ariens 2013).

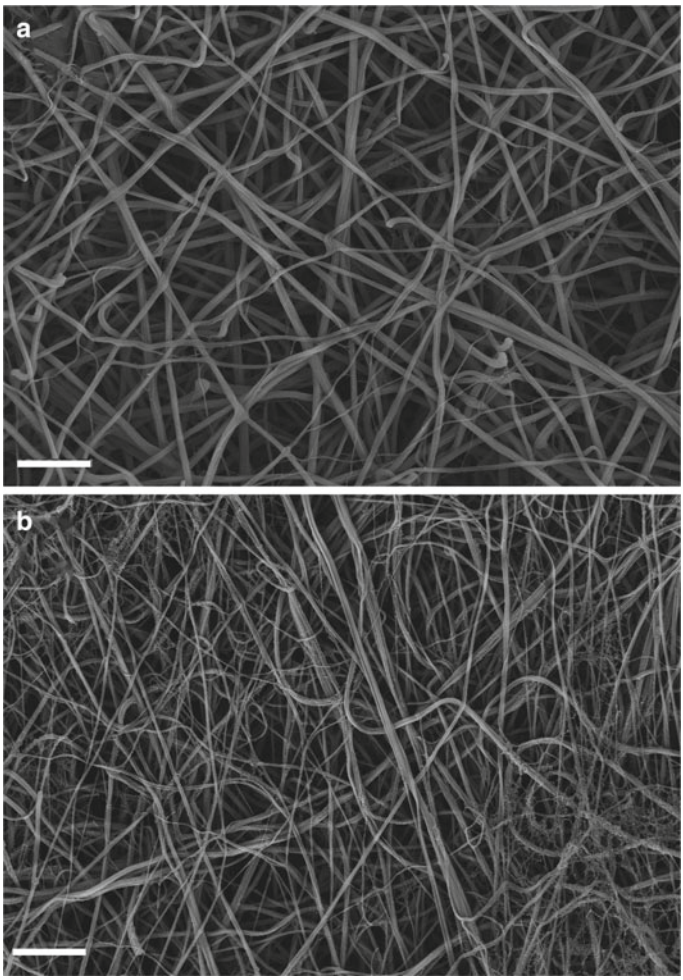


Fig. 15.10 Scanning electron microscopy image of a normal clot (A) versus a clot with a typical prothrombotic structure (B) (Weisel lab images). Magnification bars = 2 μm

In addition, a number of fibrinogen variants resulting in altered clot properties have been linked to cardiovascular disease. The most common of these is γ' fibrinogen. Fibrinogen γ' forms clots with larger pores, decreased protofibril packing within fibers, less stiff fibers, and with a heterogeneous clot structure resistant to lysis (Fig. 15.5) (Allan et al. 2012; Domingues et al. 2016). Although still inconclusive, current evidence suggest that increased levels of γ' fibrinogen are associated with arterial thrombosis and decreased levels with venous thrombosis (Macrae et al. 2016). Other fibrinogen variants resulting in altered fibrin structure and/or function with demonstrated association with disease include fibrinogen AαThr312Ala (atrial fibrillation and venous thromboembolism), BβArg448Lys (coronary artery disease

severity and stroke) and Bβ-455G/A (ischemic stroke and coronary artery disease) as well as the Factor XIIIA Val34Leu polymorphism (myocardial infarction and deep vein thrombosis) (Scott et al. 2004; Bridge et al. 2014; Chen et al. 2014). Clinically, congenital fibrinogen disorders are classified as quantitative (affecting fibrinogen concentration) or qualitative (affecting fibrinogen functionality) or a combination of both and can be asymptomatic or have a thrombotic or bleeding phenotype. For a detailed description of the clinical diagnosis and classification of congenital fibrinogen disorders please refer to Casini et al. (2018).

Recent prospective evidence of fibrin clot properties predicting disease severity and outcome provides compelling evidence that fibrin clot properties are not only associated with disease but likely play a causative role. Sumaya et al. (2018) found in a cohort of 4,345 acute coronary syndrome patients that in vitro plasma clots that were resistant to lysis, independently predicted death after adjusting for known cardiovascular risk factors at one-year follow-up. A study on 320 first-event deep vein thrombosis patients, found that reduced clot permeability (Ks) and prolonged clot lysis times independently predicted recurrent deep vein thrombosis at 44-month follow-up (Cieslik et al. 2018). Also in antiphospholipid syndrome patients (n = 126), denser fibrin networks independently predicted recurrent thromboembolic events in multivariate analysis (Celinska-Lowenhoff et al. 2018). Furthermore, in 369 unprovoked venous thromboembolism patients, those with the above-mentioned prothrombotic clot phenotype had an increased risk of cancer development within three years of follow-up (Mrozinska et al. 2019).

Also in the clinical setting, the diagnostic use of fibrin clot properties in the management of perioperative bleeding has gained traction. Standard coagulation tests that estimate coagulation factor activity such as prothrombin and partial thromboplastin times have been supplemented with methods such as thromboelastography or thromboelastometry, which measure clot formation kinetics and mechanical properties, as mentioned above (White 2018).

Fibrin Clot Properties and Bleeding

Conversely, fibrin clot structure has also been associated with bleeding. As opposed to the more prothrombotic clot phenotype described above, fibrin clots associated with bleeding are typically weaker and generally contain looser and less rigid networks of thicker fibers that are more susceptible to fibrinolysis. Bleeding resulting from a number of congenital dysfibrinogenemias and hemophilia, amongst others, have been ascribed to these clot properties (He et al. 2003; Wolberg et al. 2005; Leong et al. 2017; Casini et al. 2018). For example, the fibrin clot properties described above, have been demonstrated to distinguish severe FXI deficient patients as “bleeders” or “non-bleeders” (Zucker et al. 2014). Similarly, in hemophilia A and C, differences in fibrin clot structure are also used to assess clinical outcomes such as bleeding (Cawthern et al. 1998).

Summary

The molecular structure of fibrinogen has largely been resolved using crystallography and modeling/computation. Although certain aspects pertaining to lateral aggregation and clot growth remain to be elucidated, extensive research into fibrin polymerization has significantly enhanced our understanding of the fundamental mechanisms of clot formation and dissolution. This knowledge, together with the identification of the structural origins of the mechanical properties of fibrin has proved to be useful in various diseases with thrombotic or bleeding complications. It currently forms the basis of a number of diagnostic tools pertaining to thrombosis and/or bleeding risk and also provides a novel therapeutic target as a means to modulate blood clotting in clinical medicine. In short, fibrin is a remarkable and versatile biomaterial with a plethora of potential biological uses.

References

Allan P, Uitte de Willige S, Abou-Saleh RH, Connell SD, Ariens RA (2012) Evidence that fibrinogen gamma' directly interferes with protofibril growth: implications for fibrin structure and clot stiffness. J Thromb Haemost 10(6):1072–1080. https://doi.org/10.1111/j.1538-7836.2012.04717.x

Altieri DC, Agbanyo FR, Plescia J, Ginsberg MH, Edgington TS, Plow EF (1990) A unique recognition site mediates the interaction of fibrinogen with the leukocyte integrin Mac-1 (CD11b/CD18). J Biol Chem 265(21):12119–12122

Ariens RA (2013) Fibrin(ogen) and thrombotic disease. J Thromb Haemost 11(Suppl 1):294–305. https://doi.org/10.1111/jth.12229

Blombäck B, Hessel B, Hogg D, Therkildsen L (1978) A two-step fibrinogen–fibrin transition in blood coagulation. Nature 275(5680):501–505. https://doi.org/10.1038/275501a0

Bolliger D, Seeberger MD, Tanaka KA (2012) Principles and practice of thromboelastography in clinical coagulation management and transfusion practice. Transfus Med Rev 26(1):1–13. https://doi.org/10.1016/j.tmrv.2011.07.005

Borissoff JI, Spronk HM, ten Cate H (2011) The hemostatic system as a modulator of atherosclerosis. N Engl J Med 364(18):1746–1760. https://doi.org/10.1056/NEJMra1011670

Bowley SR, Okumura N, Lord ST (2009) Impaired protofibril formation in fibrinogen gamma N308K is due to altered D: D and "A:a" interactions. Biochemistry 48(36):8656–8663. https://doi.org/10.1021/bi900239b

Bridge KI, Philippou H, Ariens R (2014) Clot properties and cardiovascular disease. Thromb Haemost 112(5):901–908. https://doi.org/10.1160/th14-02-0184

Britton S, Kim O, Pancaldi F, Xu Z, Litvinov RI, Weisel JW, Alber M (2019) Contribution of nascent cohesive fiber-fiber interactions to the non-linear elasticity of fibrin networks under tensile load. Acta Biomater 94:514–523. https://doi.org/10.1016/j.actbio.2019.05.068

Brown JH, Volkmann N, Jun G, Henschen-Edman AH, Cohen C (2000) The crystal structure of modified bovine fibrinogen. Proc Natl Acad Sci USA 97(1):85–90. https://doi.org/10.1073/pnas.97.1.85

Brown AE, Litvinov RI, Discher DE, Purohit PK, Weisel JW (2009) Multiscale mechanics of fibrin polymer: gel stretching with protein unfolding and loss of water. Science 325(5941):741–744. https://doi.org/10.1126/science.1172484

Brown AC, Baker SR, Douglas AM, Keating M, Alvarez-Elizondo MB, Botvinick EL, Guthold M, Barker TH (2015) Molecular interference of fibrin's divalent polymerization mechanism enables

modulation of multiscale material properties. Biomaterials 49:27–36. https://doi.org/10.1016/j.biomaterials.2015.01.010

Caracciolo G, De Spirito M, Castellano AC, Pozzi D, Amiconi G, De Pascalis A, Caminiti R, Arcovito G (2003) Protofibrils within fibrin fibres are packed together in a regular array. Thromb Haemost 89(4):632–636

Casini A, Undas A, Palla R, Thachil J, de Moerloose P (2018) Diagnosis and classification of congenital fibrinogen disorders: communication from the SSC of the ISTH. J Thromb Haemost 16(9):1887–1890. https://doi.org/10.1111/jth.14216

Cawthern KM, van 't Veer C, Lock JB, DiLorenzo ME, Branda RF, Mann KG (1998) Blood coagulation in hemophilia A and hemophilia C. Blood 91(12):4581–4593

Celinska-Lowenhoff M, Zabczyk M, Iwaniec T, Plens K, Musial J, Undas A (2018) Reduced plasma fibrin clot permeability is associated with recurrent thromboembolic events in patients with antiphospholipid syndrome. Rheumatology (Oxford) 57(8):1340–1349. https://doi.org/10.1093/rheumatology/key089

Chen F, Qiao Q, Xu P, Fan B, Chen Z (2014) Effect of factor XIII-A Val34Leu polymorphism on myocardial infarction risk: a meta-analysis. Clin Appl Thromb Hemost 20(8):783–792. https://doi.org/10.1177/1076029613504130

Chernysh IN, Weisel JW (2008) Dynamic imaging of fibrin network formation correlated with other measures of polymerization. Blood 111(10):4854–4861. https://doi.org/10.1182/blood-2007-08-105247

Chernysh IN, Nagaswami C, Weisel JW (2011) Visualization and identification of the structures formed during early stages of fibrin polymerization. Blood 117(17):4609–4614. https://doi.org/10.1182/blood-2010-07-297671

Chernysh IN, Nagaswami C, Peshkova AD, Kosolapova S, Cuker A, Cines DB, Cambor CL, Litvinov RI, Weisel JW (2020) The distinctive structure and composition of arterial and venous thrombi and pulmonary emboli. Sci Rep 10(1):5112. https://doi.org/10.1038/s41598-020-59526-x

Chung DW, Davie EW (1984) Gamma and gamma' chains of human fibrinogen are produced by alternative mRNA processing. Biochemistry 23(18):4232–4236. https://doi.org/10.1021/bi00313a033

Cieslik J, Mrozinska S, Broniatowska E, Undas A (2018) Altered plasma clot properties increase the risk of recurrent deep vein thrombosis: a cohort study. Blood 131(7):797–807. https://doi.org/10.1182/blood-2017-07-798306

Cohen C, Parry DA (1990) Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. Proteins 7(1):1–15. https://doi.org/10.1002/prot.340070102

Collet JP, Montalescot G, Lesty C, Weisel JW (2002) A structural and dynamic investigation of the facilitating effect of glycoprotein IIb/IIIa inhibitors in dissolving platelet-rich clots. Circ Res 90(4):428–434. https://doi.org/10.1161/hh0402.105095

Collet J-P, Moen JL, Veklich YI, Gorkun OV, Lord ST, Montalescot G, Weisel JW (2005) The αC domains of fibrinogen affect the structure of the fibrin clot, its physical properties, and its susceptibility to fibrinolysis. Blood 106(12):3824–3830. https://doi.org/10.1182/blood-2005-05-2150

Collet JP, Allali Y, Lesty C, Tanguy ML, Silvain J, Ankri A, Blanchet B, Dumaine R, Gianetti J, Payot L, Weisel JW, Montalescot G (2006) Altered fibrin architecture is associated with hypofibrinolysis and premature coronary atherothrombosis. Arterioscler Thromb Vasc Biol 26(11):2567–2573. https://doi.org/10.1161/01.ATV.0000241589.52950.4c

De Vries JJ, Snoek CJM, Rijken DC, De Maat MPM (2020) Effects of post-translational modifications of fibrinogen on clot formation, clot structure, and fibrinolysis. Arterioscler Thromb Vasc Biol 40 (in press). https://doi.org/10.1161/atvbaha.119.313626

Domingues MM, Macrae FL, Duval C, McPherson HR, Bridge KI, Ajjan RA, Ridger VC, Connell SD, Philippou H, Ariens RA (2016) Thrombin and fibrinogen gamma' impact clot structure by marked effects on intrafibrillar structure and protofibril packing. Blood 127(4):487–495. https://doi.org/10.1182/blood-2015-06-652214

Doolittle RF, Pandi L (2006) Binding of synthetic B knobs to fibrinogen changes the character of fibrin and inhibits its ability to activate tissue plasminogen activator and its destruction by plasmin. Biochemistry 45(8):2657–2667. https://doi.org/10.1021/bi0524767

Duval C, Profumo A, Aprile A, Salis A, Millo E, Damonte G, Gauer JS, Ariëns RAS, Rocco M (2020) Fibrinogen αC-regions are not directly involved in fibrin polymerization as evidenced by a "Double-Detroit" recombinant fibrinogen mutant and knobs-mimic peptides. J Thromb Haemost 18(4):802–814. https://doi.org/10.1111/jth.14725

Erickson HP, Fowler WE (1983) Electron microscopy of fibrinogen, its plasmic fragments and small polymers. Ann N Y Acad Sci 408:146–163. https://doi.org/10.1111/j.1749-6632.1983.tb23242.x

Everse SJ, Spraggon G, Doolittle RF (1998a) A three-dimensional consideration of variant human fibrinogens. Thromb Haemost 80(1):1–9

Everse SJ, Spraggon G, Veerapandian L, Riley M, Doolittle RF (1998b) Crystal structure of fragment double-D from human fibrin with two different bound ligands. Biochemistry 37(24):8637–8642. https://doi.org/10.1021/bi9804129

Falvo MR, Millard D, O'Brien ET 3rd, Superfine R, Lord ST (2008) Length of tandem repeats in fibrin's alphaC region correlates with fiber extensibility. J Thromb Haemost 6(11):1991–1993. https://doi.org/10.1111/j.1538-7836.2008.03147.x

Flick MJ, Du X, Witte DP, Jirouskova M, Soloviev DA, Busuttil SJ, Plow EF, Degen JL (2004) Leukocyte engagement of fibrin(ogen) *via* the integrin receptor alphaMbeta2/Mac-1 is critical for host inflammatory response in vivo. J Clin Invest 113(11):1596–1606. https://doi.org/10.1172/jci20741

Fogelson AL, Keener JP (2010) Toward an understanding of fibrin branching structure. Phys Rev E: Stat, Nonlin, Soft Matter Phys 81(5):051922. https://doi.org/10.1103/PhysRevE.81.051922

Fowler WE, Erickson HP (1979) Trinodular structure of fibrinogen. Confirmation by both shadowing and negative stain electron microscopy. J Mol Biol 134(2):241–249. https://doi.org/10.1016/0022-2836(79)90034-2

Galanakis DK (1993) Inherited dysfibrinogenemia: emerging abnormal structure associations with pathologic and nonpathologic dysfunctions. Semin Thromb Hemost 19(4):386–395. https://doi.org/10.1055/s-2007-993290

Galanakis DK, Henschen A, Peerschke EI, Kehl M (1989) Fibrinogen Stony Brook, a heterozygous A alpha 16Arg—Cys dysfibrinogenemia. Evaluation of diminished platelet aggregation support and of enhanced inhibition of fibrin assembly. J Clin Invest 84(1):295–304. https://doi.org/10.1172/jci114154

Galanakis D, Spitzer S, Scharrer I (1993) Unusual A alpha 16Arg– > Cys dysfibrinogenaemic family: absence of normal A alpha-chains in fibrinogen from two of four heterozygous siblings. Blood Coagul Fibrinolysis 4(1):67–71

Ganter MT, Hofer CK (2008) Coagulation monitoring: current techniques and clinical use of viscoelastic point-of-care coagulation devices. Anesth Analg 106(5):1366–1375. https://doi.org/10.1213/ane.0b013e318168b367

Gersh KC, Nagaswami C, Weisel JW (2009a) Fibrin network structure and clot mechanical properties are altered by incorporation of erythrocytes. Thromb Haemost 102(6):1169–1175. https://doi.org/10.1160/th09-03-0199

Gersh KC, Nagaswami C, Weisel JW, Lord ST (2009b) The presence of gamma' chain impairs fibrin polymerization. Thromb Res 124(3):356–363. https://doi.org/10.1016/j.thromres.2008.11.016

Guthold M, Liu W, Sparks EA, Jawerth LM, Peng L, Falvo M, Superfine R, Hantgan RR, Lord ST (2007) A comparison of the mechanical and structural properties of fibrin fibers with other protein fibers. Cell Biochem Biophys 49(3):165–181. https://doi.org/10.1007/s12013-007-9001-4

Hall CE, Slayter HS (1959) The fibrinogen molecule: its size, shape, and mode of polymerization. J Biophys Biochem Cytol 5(1):11–16. https://doi.org/10.1083/jcb.5.1.11

Hanss M, Biot F (2001) A database for human fibrinogen variants. Ann N Y Acad Sci 936:89–90. https://doi.org/10.1111/j.1749-6632.2001.tb03495.x

He S, Blomback M, Jacobsson Ekman G, Hedner U (2003) The role of recombinant factor VIIa (FVIIa) in fibrin structure in the absence of FVIII/FIX. J Thromb Haemost 1(6):1215–1219. https://doi.org/10.1046/j.1538-7836.2003.00242.x

Henschen A, McDonagh J (1986) Chapter 7 fibrinogen, fibrin and factor XIII. In: Neuberger A, van Deenen LLM (eds) New comprehensive biochemistry, vol 13. Elsevier, pp 171–241. https://doi.org/10.1016/s0167-7306(08)60053-8

Hirota-Kawadobora M, Terasawa F, Yonekawa O, Sahara N, Shimizu E, Okumura N, Katsuyama T, Shigematsu H (2003) Fibrinogens Kosai and Ogasa: Bβ15Gly → Cys (GGT → TGT) substitution associated with impairment of fibrinopeptide B release and lateral aggregation. J Thromb Haemost 1(2):275–283. https://doi.org/10.1046/j.1538-7836.2003.00052.x

Hudson NE, Houser JR, O'Brien ET, Taylor RM, Superfine R, Lord ST, Falvo MR (2010) Stiffening of individual fibrin fibers equitably distributes strain and strengthens networks. Biophys J 98(8):1632–1640. https://doi.org/10.1016/j.bpj.2009.12.4312

Hudson NE, Ding F, Bucay I, O'Brien ET, Gorkun OV, Superfine R, Lord ST, Dokholyan NV, Falvo MR (2013) Submillisecond elastic recoil reveals molecular origins of fibrin fiber mechanics. Biophys J 104(12):2671–2680. https://doi.org/10.1016/j.bpj.2013.04.052

Hugenholtz GC, Macrae F, Adelmeijer J, Dulfer S, Porte RJ, Lisman T, Ariens RA (2016) Procoagulant changes in fibrin clot structure in patients with cirrhosis are associated with oxidative modifications of fibrinogen. J Thromb Haemost 14(5):1054–1066. https://doi.org/10.1111/jth.13278

Janmey PA, Amis EJ, Ferry JD (1983) Rheology of fibrin clots. VI. Stress relaxation, creep, and differential dynamic modulus of fine clots in large shearing deformations. J Rheol 27(2):135–153. https://doi.org/10.1122/1.549722

Jansen KA, Zhmurov A, Vos BE, Portale G, Merino DH, Litvinov RI, Tutwiler V, Kurniawan NA, Bras W, Weisel JW, Barsegov V, Koenderink GH (2020) Molecular packing structure of fibrin fibers resolved by X-ray scattering and molecular modeling. Soft Matter 16:8272–8283. https://doi.org/10.1039/d0sm00916d

Kang H, Wen Q, Janmey PA, Tang JX, Conti E, MacKintosh FC (2009) Nonlinear elasticity of stiff filament networks: strain stiffening, negative normal stress, and filament alignment in fibrin gels. J Phys Chem B 113(12):3799–3805. https://doi.org/10.1021/jp807749f

Kattula S, Byrnes JR, Wolberg AS (2017) Fibrinogen and fibrin in hemostasis and thrombosis. Arterioscler Thromb Vasc Biol 37(3):e13–e21. https://doi.org/10.1161/atvbaha.117.308564

Kim OV, Litvinov RI, Weisel JW, Alber MS (2014) Structural basis for the nonlinear mechanics of fibrin networks under compression. Biomaterials 35(25):6739–6749. https://doi.org/10.1016/j.biomaterials.2014.04.056

Kim OV, Liang X, Litvinov RI, Weisel JW, Alber MS, Purohit PK (2016) Foam-like compression behavior of fibrin networks. Biomech Model Mechanobiol 15(1):213–228. https://doi.org/10.1007/s10237-015-0683-z

Klykov O, van der Zwaan C, Heck AJR, Meijer AB, Scheltema RA (2020) Missing regions within the molecular architecture of human fibrin clots structurally resolved by XL-MS and integrative structural modeling. Proc Natl Acad Sci U S A 117(4):1976–1987. https://doi.org/10.1073/pnas.1911785117

Köhler S, Schmid F, Settanni G (2015) The internal dynamics of fibrinogen and its implications for coagulation and adsorption. PLoS Comput Biol 11(9):e1004346–e1004346. https://doi.org/10.1371/journal.pcbi.1004346

Kollman JM, Pandi L, Sawaya MR, Riley M, Doolittle RF (2009) Crystal structure of human fibrinogen. Biochemistry 48(18):3877–3886. https://doi.org/10.1021/bi802205g

Kostelansky MS, Betts L, Gorkun OV, Lord ST (2002) 2.8 A crystal structures of recombinant fibrinogen fragment D with and without two peptide ligands: GHRP binding to the "b" site disrupts its nearby calcium-binding site. Biochemistry 41(40):12124–12132. https://doi.org/10.1021/bi0261894

Langer BG, Weisel JW, Dinauer PA, Nagaswami C, Bell WR (1988) Deglycosylation of fibrinogen accelerates polymerization and increases lateral aggregation of fibrin fibers. J Biol Chem 263(29):15056–15063

Leong L, Chernysh IN, Xu Y, Sim D, Nagaswami C, de Lange Z, Kosolapova S, Cuker A, Kauser K, Weisel JW (2017) Clot stability as a determinant of effective factor VIII replacement in hemophilia A. Res Pract Thromb Haemost 1(2):231–241. https://doi.org/10.1002/rth2.12034

Lepedda AJ, Cigliano A, Cherchi GM, Spirito R, Maggioni M, Carta F, Turrini F, Edelstein C, Scanu AM, Formato M (2009) A proteomic approach to differentiate histologically classified stable and unstable plaques from human carotid arteries. Atherosclerosis 203(1):112–118. https://doi.org/10.1016/j.atherosclerosis.2008.07.001

Li W, Sigley J, Pieters M, Helms CC, Nagaswami C, Weisel JW, Guthold M (2016) Fibrin fiber stiffness is strongly affected by fiber diameter, but not by fibrinogen glycation. Biophys J 110(6):1400–1410. https://doi.org/10.1016/j.bpj.2016.02.021

Li W, Sigley J, Baker SR, Helms CC, Kinney MT, Pieters M, Brubaker PH, Cubcciotti R, Guthold M (2017) Nonuniform internal structure of fibrin fibers: Protein density and bond density strongly decrease with increasing diameter. Biomed Res Int 2017:6385628–6385628. https://doi.org/10.1155/2017/6385628

Liang X, Chernysh I, Purohit PK, Weisel JW (2017) Phase transitions during compression and decompression of clots from platelet-poor plasma, platelet-rich plasma and whole blood. Acta Biomater 60:275–290. https://doi.org/10.1016/j.actbio.2017.07.011

Lisman T, Ariens RA (2016) Alterations in fibrin structure in patients with liver diseases. Semin Thromb Hemost 42(4):389–396. https://doi.org/10.1055/s-0036-1572327

Litvinov RI, Weisel JW (2016) What is the biological and clinical relevance of fibrin? Semin Thromb Hemost 42(4):333–343. https://doi.org/10.1055/s-0036-1571342

Litvinov RI, Gorkun OV, Owen SF, Shuman H, Weisel JW (2005) Polymerization of fibrin: specificity, strength, and stability of knob-hole interactions studied at the single-molecule level. Blood 106(9):2944–2951. https://doi.org/10.1182/blood-2005-05-2039

Litvinov RI, Gorkun OV, Galanakis DK, Yakovlev S, Medved L, Shuman H, Weisel JW (2006) Polymerization of fibrin: direct observation and quantification of individual B:b knob-hole interactions. Blood 109(1):130–138. https://doi.org/10.1182/blood-2006-07-033910

Litvinov RI, Yakovlev S, Tsurupa G, Gorkun OV, Medved L, Weisel JW (2007) Direct evidence for specific interactions of the fibrinogen alphaC-domains with the central E region and with each other. Biochemistry 46(31):9133–9142. https://doi.org/10.1021/bi700944j

Litvinov RI, Faizullin DA, Zuev YF, Weisel JW (2012) The α-helix to β-sheet transition in stretched and compressed hydrated fibrin clots. Biophys J 103(5):1020–1027. https://doi.org/10.1016/j.bpj.2012.07.046

Litvinov RI, Kononova O, Zhmurov A, Marx KA, Barsegov V, Thirumalai D, Weisel JW (2018) Regulatory element in fibrin triggers tension-activated transition from catch to slip bonds. Proc Natl Acad Sci U S A 115(34):8575–8580. https://doi.org/10.1073/pnas.1802576115

Litvinov RI, Nabiullina RM, Zubairova LD, Shakurova MA, Andrianova IA, Weisel JW (2019) Lytic susceptibility, structure, and mechanical properties of fibrin in systemic lupus erythematosus. Front Immunol 10:1626. https://doi.org/10.3389/fimmu.2019.01626

Liu W, Jawerth LM, Sparks EA, Falvo MR, Hantgan RR, Superfine R, Lord ST, Guthold M (2006) Fibrin fibers have extraordinary extensibility and elasticity. Science 313(5787):634. https://doi.org/10.1126/science.1127317

Liu W, Carlisle CR, Sparks EA, Guthold M (2010) The mechanical properties of single fibrin fibers. J Thromb Haemost 8(5):1030–1036. https://doi.org/10.1111/j.1538-7836.2010.03745.x

Longstaff C, Varju I, Sotonyi P, Szabo L, Krumrey M, Hoell A, Bota A, Varga Z, Komorowicz E, Kolev K (2013) Mechanical stability and fibrinolytic resistance of clots containing fibrin, DNA, and histones. J Biol Chem 288(10):6946–6956. https://doi.org/10.1074/jbc.M112.404301

Luddington RJ (2005) Thrombelastography/thromboelastometry. Clin Lab Haematol 27(2):81–90. https://doi.org/10.1111/j.1365-2257.2005.00681.x

Macrae FL, Domingues MM, Casini A, Ariens RA (2016) The (patho)physiology of fibrinogen gamma'. Semin Thromb Hemost 42(4):344–355. https://doi.org/10.1055/s-0036-1572353

Madrazo J, Brown JH, Litvinovich S, Dominguez R, Yakovlev S, Medved L, Cohen C (2001) Crystal structure of the central region of bovine fibrinogen (E5 fragment) at 1.4-A resolution. Proc Natl Acad Sci USA 98(21):11967–11972. https://doi.org/10.1073/pnas.211439798

Marchi RC, Carvajal Z, Boyer-Neumann C, Anglés-Cano E, Weisel JW (2006) Functional characterization of fibrinogen Bicêtre II: a gamma 308 Asn– > Lys mutation located near the fibrin D: D interaction sites. Blood Coagul Fibrinolysis 17(3):193–201. https://doi.org/10.1097/01.mbc.0000220241.22714.68

Marsh JJ, Guan HS, Li S, Chiles PG, Tran D, Morris TA (2013) Structural insights into fibrinogen dynamics using amide hydrogen/deuterium exchange mass spectrometry. Biochemistry 52(32):5491–5502. https://doi.org/10.1021/bi4007995

Matsuka YV, Medved LV, Migliorini MM, Ingham KC (1996) Factor XIIIa-catalyzed cross-Linking of recombinant αC fragments of human fibrinogen. Biochemistry 35(18):5810–5816. https://doi.org/10.1021/bi952294k

Medved L, Weisel JW (2009) Recommendations for nomenclature on fibrinogen and fibrin. J Thromb Haemost 7(2):355–359. https://doi.org/10.1111/j.1538-7836.2008.03242.x

Medved L, Ugarova T, Veklich Y, Lukinova N, Weisel J (1990) Electron microscope investigation of the early stages of fibrin assembly. Twisted protofibrils and fibers. J Mol Biol 216(3):503–509. https://doi.org/10.1016/0022-2836(90)90376-w

Mills JD, Ariens RA, Mansfield MW, Grant PJ (2002) Altered fibrin clot structure in the healthy relatives of patients with premature coronary artery disease. Circulation 106(15):1938–1942. https://doi.org/10.1161/01.cir.0000033221.73082.06

Missori M, Papi M, Maulucci G, Arcovito G, Boumis G, Bellelli A, Amiconi G, De Spirito M (2010) Cl—and F—anions regulate the architecture of protofibrils in fibrin gel. Eur Biophys J 39(6):1001–1006. https://doi.org/10.1007/s00249-009-0492-3

Mosesson MW, Finlayson JS, Umfleet RA (1972) Human fibrinogen heterogeneities. 3. Identification of chain variants. J Biol Chem 247(16):5223–5227

Mosesson MW, DiOrio JP, Siebenlist KR, Wall JS, Hainfeld JF (1993) Evidence for a second type of fibril branch point in fibrin polymer networks, the trimolecular junction. Blood 82(5):1517–1521

Mrozinska S, Cieslik J, Broniatowska E, Malinowski KP, Undas A (2019) Prothrombotic fibrin clot properties associated with increased endogenous thrombin potential and soluble P-selectin predict occult cancer after unprovoked venous thromboembolism. J Thromb Haemost 17(11):1912–1922. https://doi.org/10.1111/jth.14579

Mullin JL, Gorkun OV, Lord ST (2000) Decreased lateral aggregation of a variant recombinant fibrinogen provides insight into the polymerization mechanism. Biochemistry 39(32):9843–9849. https://doi.org/10.1021/bi000045c

O'Brien ET, Falvo MR, Millard D, Eastwood B, Taylor RM, Superfine R (2008) Ultrathin self-assembled fibrin sheets. Proc Natl Acad Sci U S A 105(49):19438–19443. https://doi.org/10.1073/pnas.0804865105

Okumura N, Gorkun OV, Lord ST (1997) Severely impaired polymerization of recombinant fibrinogen gamma-364 Asp – > His, the substitution discovered in a heterozygous individual. J Biol Chem 272(47):29596–29601. https://doi.org/10.1074/jbc.272.47.29596

Okumura N, Terasawa F, Hirota-Kawadobora M, Yamauchi K, Nakanishi K, Shiga S, Ichiyama S, Saito M, Kawai M, Nakahata T (2006) A novel variant fibrinogen, deletion of Bβ111Ser in coiled-coil region, affecting fibrin lateral aggregation. Clin Chim Acta 365(1):160–167. https://doi.org/10.1016/j.cca.2005.08.014

Okumura N, Terasawa F, Haneishi A, Fujihara N, Hirota-Kawadobora M, Yamauchi K, Ota H, Lord ST (2007) B:b interactions are essential for polymerization of variant fibrinogens with impaired holes 'a'1. J Thromb Haemost 5(12):2352–2359. https://doi.org/10.1111/j.1538-7836.2007.02793.x

Palka I, Nessler J, Nessler B, Piwowarska W, Tracz W, Undas A (2010) Altered fibrin clot properties in patients with chronic heart failure and sinus rhythm: a novel prothrombotic mechanism. Heart 96(14):1114–1118. https://doi.org/10.1136/hrt.2010.192740

Piechocka IK, Bacabac RG, Potters M, Mackintosh FC, Koenderink GH (2010) Structural hierarchy governs fibrin gel mechanics. Biophys J 98(10):2281–2289. https://doi.org/10.1016/j.bpj.2010.01.040

Pieters M, Wolberg AS (2019) Fibrinogen and fibrin: An illustrated review. Res Pract Thromb Haemost 3(2):161–172. https://doi.org/10.1002/rth2.12191

Ping L, Huang L, Cardinali B, Profumo A, Gorkun OV, Lord ST (2011) Substitution of the human αC region with the analogous chicken domain generates a fibrinogen with severely impaired lateral aggregation: fibrin monomers assemble into protofibrils but protofibrils do not assemble into fibers. Biochemistry 50(42):9066–9075. https://doi.org/10.1021/bi201094v

Protopopova AD, Barinov NA, Zavyalova EG, Kopylov AM, Sergienko VI, Klinov DV (2015) Visualization of fibrinogen αC regions and their arrangement during fibrin network formation by high-resolution AFM. J Thromb Haemost 13(4):570–579. https://doi.org/10.1111/jth.12785

Protopopova AD, Litvinov RI, Galanakis DK, Nagaswami C, Barinov NA, Mukhitov AR, Klinov DV, Weisel JW (2017) Morphometric characterization of fibrinogen's αC regions and their role in fibrin self-assembly and molecular organization. Nanoscale 9(36):13707–13716. https://doi.org/10.1039/c7nr04413e

Purohit PK, Litvinov RI, Brown AEX, Discher DE, Weisel JW (2011) Protein unfolding accounts for the unusual mechanical behavior of fibrin networks. Acta Biomater 7(6):2374–2383. https://doi.org/10.1016/j.actbio.2011.02.026

Riha P, Wang X, Liao R, Stoltz JF (1999) Elasticity and fracture strain of whole blood clots. Clin Hemorheol Microcirc 21(1):45–49

Rocco M, Molteni M, Ponassi M, Giachi G, Frediani M, Koutsioubas A, Profumo A, Trevarin D, Cardinali B, Vachette P, Ferri F, Pérez J (2014) A comprehensive mechanism of fibrin network formation involving early branching and delayed single-to double-strand transition from coupled time-resolved X-ray/light-scattering detection. J Am Chem Soc 136(14):5376–5384. https://doi.org/10.1021/ja5002955

Ryan EA, Mockros LF, Weisel JW, Lorand L (1999) Structural origins of fibrin clot rheology. Biophys J 77(5):2813–2826. https://doi.org/10.1016/S0006-3495(99)77113-4

Scott EM, Ariens RA, Grant PJ (2004) Genetic and environmental determinants of fibrin structure and function: relevance to clinical disease. Arterioscler Thromb Vasc Biol 24(9):1558–1566. https://doi.org/10.1161/01.ATV.0000136649.83297.bf

Spraggon G, Everse SJ, Doolittle RF (1997) Crystal structures of fragment D from human fibrinogen and its crosslinked counterpart from fibrin. Nature 389(6650):455–462. https://doi.org/10.1038/38947

Standeven KF, Carter AM, Grant PJ, Weisel JW, Chernysh I, Masova L, Lord ST, Ariëns RAS (2007) Functional analysis of fibrin γ-chain cross-linking by activated factor XIII: determination of a cross-linking pattern that maximizes clot stiffness. Blood 110(3):902–907. https://doi.org/10.1182/blood-2007-01-066837

Storm C, Pastore JJ, MacKintosh FC, Lubensky TC, Janmey PA (2005) Nonlinear elasticity in biological gels. Nature 435(7039):191–194. https://doi.org/10.1038/nature03521

Sumaya W, Wallentin L, James SK, Siegbahn A, Gabrysch K, Bertilsson M, Himmelmann A, Ajjan RA, Storey RF (2018) Fibrin clot properties independently predict adverse clinical outcome following acute coronary syndrome: a PLATO substudy. Eur Heart J 39(13):1078–1085. https://doi.org/10.1093/eurheartj/ehy013

Torbet J, Freyssinet JM, Hudry-Clergeon G (1981) Oriented fibrin gels formed by polymerization in strong magnetic fields. Nature 289(5793):91–93. https://doi.org/10.1038/289091a0

Trezzini C, Jungi TW, Kuhnert P, Peterhans E (1988) Fibrinogen association with human monocytes: evidence for constitutive expression of fibrinogen receptors and for involvement of Mac-1 (CD18, CR3) in the binding. Biochem Biophys Res Commun 156(1):477–484. https://doi.org/10.1016/s0006-291x(88)80866-0

Tsurupa G, Mahid A, Veklich Y, Weisel JW, Medved L (2011) Structure, stability, and interaction of fibrin αC-domain polymers. Biochemistry 50(37):8028–8037. https://doi.org/10.1021/bi2008189
Undas A (2016) How to assess fibrinogen levels and fibrin clot properties in clinical practice? Semin Thromb Hemost 42(4):381–388. https://doi.org/10.1055/s-0036-1579636
Undas A, Ariens RA (2011) Fibrin clot structure and function: a role in the pathophysiology of arterial and venous thromboembolic diseases. Arterioscler Thromb Vasc Biol 31(12):e88–99. https://doi.org/10.1161/atvbaha.111.230631
Undas A, Szuldrzynski K, Stepien E, Zalewski J, Godlewski J, Tracz W, Pasowicz M, Zmudka K (2008) Reduced clot permeability and susceptibility to lysis in patients with acute coronary syndrome: effects of inflammation and oxidative stress. Atherosclerosis 196(2):551–557. https://doi.org/10.1016/j.atherosclerosis.2007.05.028
Undas A, Slowik A, Wolkow P, Szczudlik A, Tracz W (2010) Fibrin clot properties in acute ischemic stroke: relation to neurological deficit. Thromb Res 125(4):357–361. https://doi.org/10.1016/j.thromres.2009.11.013
Veklich YI, Gorkun OV, Medved LV, Nieuwenhuizen W, Weisel JW (1993) Carboxyl-terminal portions of the alpha chains of fibrinogen and fibrin. Localization by electron microscopy and the effects of isolated alpha C fragments on polymerization. J Biol Chem 268(18):13577–13585
Vos BE, Martinez-Torres C, Burla F, Weisel JW, Koenderink GH (2020) Revealing the molecular origins of fibrin's elastomeric properties by in situ X-ray scattering. Acta Biomater 104:39–52. https://doi.org/10.1016/j.actbio.2020.01.002
Weisel JW (1986) The electron microscope band pattern of human fibrin: various stains, lateral order, and carbohydrate localization. J Ultrastruct Mol Struct Res 96(1–3):176–188. https://doi.org/10.1016/0889-1605(86)90019-4
Weisel JW (2004) The mechanical properties of fibrin for basic scientists and clinicians. Biophys Chem 112(2–3):267–276. https://doi.org/10.1016/j.bpc.2004.07.029
Weisel JW, Litvinov RI (2013) Mechanisms of fibrin polymerization and clinical implications. Blood 121(10):1712–1719. https://doi.org/10.1182/blood-2012-09-306639
Weisel JW, Litvinov RI (2017) Fibrin formation, structure and properties. Sub-cellular biochemistry 82:405–456. https://doi.org/10.1007/978-3-319-49674-0_13
Weisel JW, Medved L (2001) The structure and function of the alpha C domains of fibrinogen. Ann NY Acad Sci 936:312–327. https://doi.org/10.1111/j.1749-6632.2001.tb03517.x
Weisel JW, Warren SG, Cohen C (1978) Crystals of modified fibrinogen: size, shape and packing of molecules. J Mol Biol 126(2):159–183. https://doi.org/10.1016/0022-2836(78)90357-1
Weisel JW, Phillips GN Jr, Cohen C (1983) The structure of fibrinogen and fibrin: II. Architecture of the fibrin clot. Ann N Y Acad Sci 408:367–379. https://doi.org/10.1111/j.1749-6632.1983.tb23257.x
Weisel JW, Stauffacher CV, Bullitt E, Cohen C (1985) A model for fibrinogen: domains and sequence. Science 230(4732):1388–1391. https://doi.org/10.1126/science.4071058
White NJ (2018) Clot structure predicts recurrent thrombosis. Blood 131(7):715–716. https://doi.org/10.1182/blood-2018-01-824169
Williams RC (1981) Morphology of bovine fibrinogen monomers and fibrin oligomers. J Mol Biol 150(3):399–408. https://doi.org/10.1016/0022-2836(81)90555-6
Wolberg AS, Allen GA, Monroe DM, Hedner U, Roberts HR, Hoffman M (2005) High dose factor VIIa improves clot structure and stability in a model of haemophilia B. Br J Haematol 131(5):645–655. https://doi.org/10.1111/j.1365-2141.2005.05820.x
Wolfenstein-Todel C, Mosesson MW (1981) Carboxy-terminal amino acid sequence of a human fibrinogen gamma-chain variant (gamma'). Biochemistry 20(21):6146–6149. https://doi.org/10.1021/bi00524a036
Yang Z, Mochalkin I, Doolittle RF (2000) A model of fibrin formation based on crystal structures of fibrinogen and fibrin fragments complexed with synthetic peptides. Proc Natl Acad Sci U S A 97(26):14156–14161. https://doi.org/10.1073/pnas.97.26.14156

Yee VC, Pratt KP, Côté HC, Trong IL, Chung DW, Davie EW, Stenkamp RE, Teller DC (1997) Crystal structure of a 30 kDa C-terminal fragment from the gamma chain of human fibrinogen. Structure 5(1):125–138. https://doi.org/10.1016/s0969-2126(97)00171-8

Yermolenko IS, Lishko VK, Ugarova TP, Magonov SN (2011) High-resolution visualization of fibrinogen molecules and fibrin fibers with atomic force microscopy. Biomacromol 12(2):370–379. https://doi.org/10.1021/bm101122g

Yeromonahos C, Polack B, Caton F (2010) Nanostructure of the fibrin clot. Biophys J 99(7):2018–2027. https://doi.org/10.1016/j.bpj.2010.04.059

Zhmurov A, Brown AE, Litvinov RI, Dima RI, Weisel JW, Barsegov V (2011) Mechanism of fibrin(ogen) forced unfolding. Structure 19(11):1615–1624. https://doi.org/10.1016/j.str.2011.08.013

Zhmurov A, Protopopova AD, Litvinov RI, Zhukov P, Mukhitov AR, Weisel JW, Barsegov V (2016) Structural basis of interfacial flexibility in fibrin oligomers. Structure 24(11):1907–1917. https://doi.org/10.1016/j.str.2016.08.009

Zhmurov A, Protopopova AD, Litvinov RI, Zhukov P, Weisel JW, Barsegov V (2018) Atomic structural models of fibrin oligomers. Structure 26(6):857–868.e854. https://doi.org/10.1016/j.str.2018.04.005

Zucker M, Seligsohn U, Salomon O, Wolberg AS (2014) Abnormal plasma clot structure and stability distinguish bleeding risk in patients with severe factor XI deficiency. J Thromb Haemost 12(7):1121–1130. https://doi.org/10.1111/jth.12600

Zuev YF, Litvinov RI, Sitnitsky AE, Idiyatullin BZ, Bakirova DR, Galanakis DK, Zhmurov A, Barsegov V, Weisel JW (2017) Conformational flexibility and self-association of fibrinogen in concentrated solutions. J Phys Chem B 121(33):7833–7843. https://doi.org/10.1021/acs.jpcb.7b05654

Chapter 16
Structural Organization and Protein-Protein Interactions in Human Adenovirus Capsid

Vijay S. Reddy and Michael A. Barry

Abstract Human adenoviruses (HAdVs) are large (150 MDa), complex, nonenveloped dsDNA viruses that cause self-limiting respiratory, ocular and enteric infections. They are significant health hazard in young, elderly and immuno-compromised populations. Moreover, various adenoviruses (AdVs) of mammalian origin are being used as vectors in gene, vaccine and cancer therapies. Multiple copies of at least 13 different proteins, all in all ~2800 protein molecules, come together to form an adenovirus virion packaging the ~36 Kbp geome. The details of structural organization of the adenovirus capsid and underlying network of protein-protein interactions provide clues into designing the modified and novel adenovirus vectors with desired functionalities and/or targeting specificities. The advancements in 3D structure determination by cryo-electron microscopy (cryo-EM) in the past decade have enabled unveiling of the complex organization of adenovirus architecture at near atomic resolution. Specifically, these studies revealed the structures and the network of interactions involving cement/minor proteins in stabilizing the AdV icosahedral architecture, which appear to be mostly conserved among human adenoviruses. In this chapter, we describe the current state of knowledge on the structure and organization of human adenoviruses.

Keywords Adenovirus structure · Protein network · Protein-protein interactions · Cement proteins · Hexon · Penton base · Protein IX · Protein IIIa · Protein VIII and protein VI

V. S. Reddy (✉)
Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
e-mail: reddyv@scripps.edu

M. A. Barry
Department of Internal Medicine, Division of Infectious Diseases, Mayo Clinic, Rochester, MN 55902, USA
e-mail: Barry.Michael@mayo.edu

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_16

Introduction

Human adenovirus capsids are large, ~100 nm in diameter, and complex assemblies composed of hundreds of copies of seven structural proteins (Table 16.1) (van Oostrum and Burnett 1985; Benevento et al. 2014). In addition to the capsid forming proteins, there are a number of core proteins including viral protease that interact with and condense the packaged dsDNA genome as well as scaffolding proteins that aid in the assembly of the adenovirus virion. Table 16.1 lists various proteins, their copy number and location in the adenovirus capsid. Figure 16.1 shows the overall geometric shape and the location and organization of various proteins present in the

Table 16.1 Various proteins and their copy number present in an adenovirus particle[a]

Polypeptide	~MW (kDa)[a]	Location	Copy number[b]
II (Hexon)	107.9	Coat	720
III (Penton base)	63.3	Coat (vertex)	60
IV (Fiber)	62.0[c]	Coat (vertex)	36
IIIa	65.2	Cement (interior)	60
VIII	24.7	Cement (interior)	120
VI	27.0	Cement (interior)	360–480
IX	14.4	Cement (exterior)	240
VII	22	Core (interior)	833
V	41.3	Core (interior)	~157
TP	74.6	Core (interior)	2
μ	8.8	Core (interior)	104
AVP	23.1	Core (interior)	~15
IVa2	50.9	Core (interior)	~7

[a]Based on human adenovirus 5
[b]Consensus copy number
[c]MW of fibers from other HAdV serotypes vary depending on the number of fiber shaft repeats they contain. The MW of the fibers from other prominent/selected AdV serotypes are Ad6 (55.8), Ad11 (35.5) Ad26 (41.0), Ad35 (35.3), Ad40 short (41.4), Ad40 long (59.1)

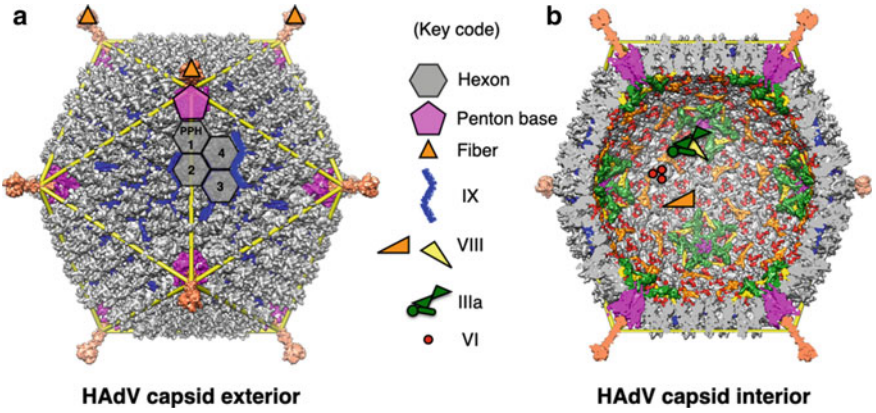


Fig. 16.1 Schematic of the structural organization of HAdV. **a** Exterior view of the HAdV organization. Different proteins that form part of the AdV capsid are shown in different colors and identified by different shapes and colors according to the key code shown on the right. The four structurally unique hexons and their positions are identified by (gray) hexagons numbered 1–4. The hexon-1 (HX-1) is commonly referred to as peripentonal hexon (PPH) that along with its 5-fold related symmetry mates surround penton bases (PBs), which are located at the icosahedral 5-fold vertices, shown in magenta color and identified by pentagons. The trimeric fiber molecules attached to the PBs are identified by orange equilateral triangles. The blue regions/lines correspond to protein IX molecule(s), which is the only the minor/cement protein located on the capsid exterior. The icosahedral cage is shown as yellow tubes. **b** A cut open view of the HAdV virion showing the interior minor proteins, IIIa, VI and VIII indicated with the corresponding shapes and colors asindicated in the key code

virion coat of a human adenovirus. Overall, all the known adenoviruses display icosahedral architecture with the 3 major capsid proteins (MCP), hexon, penton base (PB) and fiber arranged in a similar manner. Specifically, the most abundant hexon protein, present in 720 copies, is organized as trimers in pseudo T = 25 icosahedral arrangement with the pentameric PB subunits located at the icosahedral vertices (Fig. 16.1) (Stewart et al. 1991; San Martin 2012). The trimeric fiber molecules that emanate from the icosahedral vertices, interact with the pentameric PB capsomers. The interactions between the PB and fiber molecules are unique because of the mismatch of symmetries, 5-fold versus 3-fold, of the molecules involved, respectively. In addition, four different minor proteins that are present in multiple copies on the capsid exterior (IX) and interior (IIIa, VI, VIII) stabilize the icosahedral shell, primarily through interactions with the hexon subunits as detailed in the following sections.

Structures and Interactions of Major Capsid Proteins

Hexon

The hexon subunit is the largest (~110 kDa; ~950 a.a) and most abundant capsid protein of the adenovirus capsid, present in 720 copies (Stewart et al. 1991; San Martin 2012). Each hexon subunit is composed of two tandem jellyroll β-barrels (termed V1, V2) interspersed with large loop insertions (Fig. 16.2a) (Rux and Burnett 2000). The characteristic feature of hexon subunits is that they form stable and intertwined trimers (referred simply as hexons) with a lefthanded twist that exhibit perfect 3-fold symmetry (Fig. 16.2b, c). In other words, each hexon subunit wraps around its clockwise related neighbor. Thus, 720 hexon subunits are organized as 240 hexon trimers in all the structurally characterized AdV capsids. The VIPERdb analysis (Carrillo-Tripp et al. 2009) of human adenovirus 26 (HAdV-26) structure (Yu et al. 2017) suggests that a total of ~18,618 Å^2 surface area buried between a pair of hexon subunits in forming trimeric hexon, i.e., in "*intra*"-hexon associations.

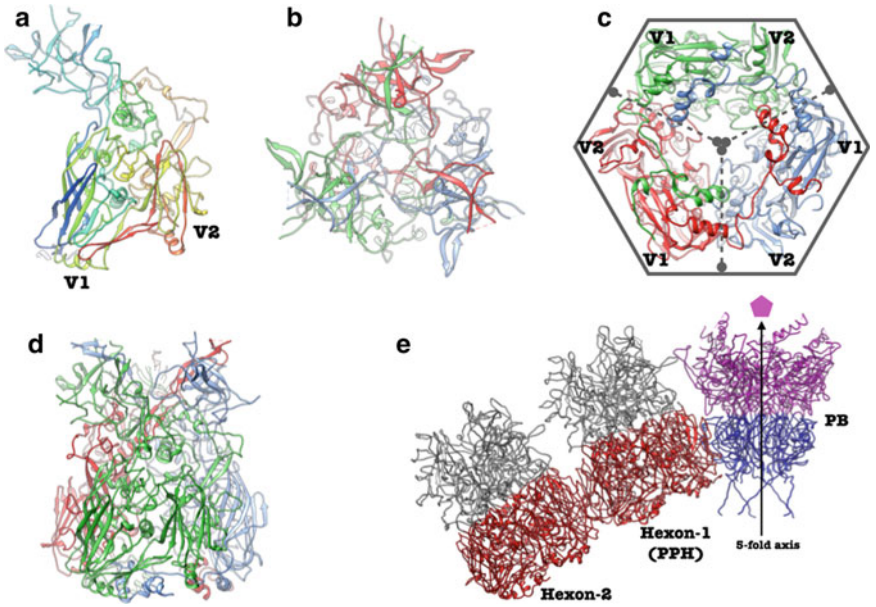


Fig. 16.2 Structure and associations of hexon subunits. **a** Tertiary structure of the hexon subunit of HAdV-C5 (PDB-ID:6B1T), the two jellyroll β-barrels are labeled V1 and V2. **b** The top view of trimeric hexon showing the triangular shape. **c** The bottom view of the hexon showing the hexagonal arrangement of the six β-barrels, two (V1 and V2) from each subunit, delineated by the dashed lines. **d** The side view of the intertwined trimeric hexon. The individual subunits are shown in different colors. **e** Association of the hexons through interactions between the β-barrels, colored red and between the β-barrels of penton (PB), colored blue, which is located at the icosahedral 5-fold vertices, identified by the magenta colored pentagon

Three times that is buried (~55,855 Å^2) in a hexon trimer. Of note, the total buried surface area (BSA) indicated above is twice the amount of contact surface area between the interacting subunits. This amounts to ~9.8 Å^2 BSA per a.a. of hexon, considering 952 a.a. in a hexon subunit and $2 \times 952 = 1904$ residues in a pair of hexon subunits (18,618/1904 = 9.77). If one considers the per a.a. BSA as a measure of strength of association between the subunits, that number is only a meagre 0.75 Å^2 BSA per a.a. on average for "*inter*"-hexon associations, which are ~1/13th of the *intra*-hexon associations. Alternatively, you can consider it as, on an average, 13 times as many residues interacting at an *intra*-hexon interface compared to an *inter*-hexon interface. Hence, the *intra*-hexon interactions are likely the strongest protein-protein interactions found in the adenovirus capsid. This in turn suggests that the trimeric hexon forms a strongest building block of the AdV capsid. This observation is supported by the isolation and purification of intact hexons by disassembling the AdV virions (Rux and Burnett 2007).

It is noteworthy that the 3D structure and trimeric organization of isolated hexon, determined by X-ray crystallography (Rux and Burnett 2000) is very similar to the hexon in the assembled virion, determined by cryo-EM and X-ray diffraction methods at near atomic resolution (Liu et al. 2010; Reddy et al. 2010; Yu et al. 2017; Dai et al. 2017; Kundhavai Natchiar et al. 2018), with the exception of ordered hyper variable regions (HVRs) that were better ordered in the context of virion due to their interactions with the neighboring subunits in the assembled particle. Remarkably, while the top view of the hexon resembles trimeric assembly (Fig. 16.2c), the bottom view assumes hexagonal shape because of the above mentioned 3-fold symmetry applied to dual β-barrels (Fig. 16.2d). The hexons associate with the neighboring hexons primarily through interactions between the β-barrels (Fig. 16.2e). The peripentonal hexons (labeled in PPH/Hexon-1, Fig. 16.1a) that surround the penton base (PB) subunits at the icosahedral 5-fold vertices also interact with the β-barrels of PBs. In addition, the hexon-hexon interactions are further "cemented" by the minor proteins IX on the capsid exterior and IIIa and VIII on the interior of the capsid (see below).

Penton Base (PB)

The characteristic feature of PB subunits is that they are organized as pentamers and occupy the icosahedral vertices of the AdV virion. The monomeric PB subunit (~63 kDa) displays a jellyroll β-barrel base with large insertions (118–383, 417–468) between the β-strands that form elaborate surface loops (Fig. 16.3a). Together, these surface loops from five PB-subunits form a shallow trough shaped depression that acts as the pedestal for the trimeric fiber molecule (Fig. 16.3b, c). Similar to hexon, the tertiary structure and pentameric (quaternary) organization of PB are well conserved in the isolated and in situ structures (Zubieta et al. 2005; Liu et al. 2010; Yu et al. 2017) (Fig. 16.3). The N-terminal residues (1–36) and the residues in the RGD loop (297–376) of Ad5 PB are disordered. While a total of ~36,190 Å^2 surface area is buried in a pentameric capsomere, a 1/5th of that (7238 Å^2) is buried between

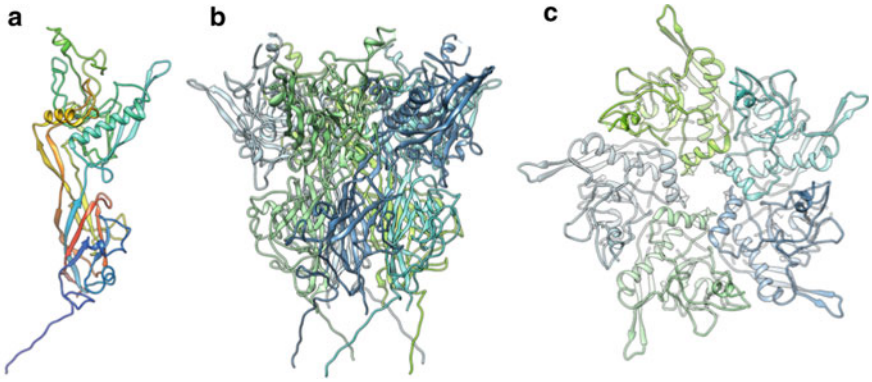


Fig. 16.3 Structure and association of penton base (PB) subunits (PDB-ID: 6B1T). **a** Tertiary fold of the monomeric PB subunit color coded in rainbow spectrum from N-terminus (blue) and C-terminus (red). **b** Side view of the PB-pentamer. Individual subunits are color coded in different colors. **c** Top view of the PB-pentamer in panel B

a pair of PB subunits, which amounts to ~7.0 Å^2 surface area buried per a.a. of PB subunit on average (7238/[2 × 520] = 6.9), which is comparable to but smaller than the *intra*-hexon interactions (~9.8 Å^2). This in turn suggests that the pentamer of PB is also a strong preassembled structure and represents another building block of AdV structure. Of note, the per a.a. BSA estimate for PB-PPH interactions is 1.1 Å^2, which is roughly similar to inter-hexon interactions (0.75 Å^2).

Fiber and PB Interactions

One of the unique protein-protein interactions found in the AdV capsid involve fiber and PB, where three molecules of the (trimeric) fiber interacts with the five molecules of (pentameric) PB, located at each of the twelve icosahedral vertices in the AdV virion. This results in 36 molecules of single kind of fiber (62 kDa) found in an AdV capsid. However, certain adenoviruses are known to have two different kinds of fiber molecules (one short and one long) (Lenman et al. 2015). The high-resolution structural details of fiber molecule were revealed from the numerous crystal structures of fiber knobs and truncated version of the fibers (Xia et al. 1994; van Raaij et al. 1999; Burmeister et al. 2004) (Fig. 16.4a, b). However, it has not been possible to determine the fiber structure in the context of AdV virion in greater detail. This lack of specifics on PB-fiber interactions is due to imposition of the 5-fold symmetry on the 3-fold symmetric fiber during the icosahedral particle reconstruction and also because of inherent flexibility of the fiber molecule. However, in conjunction with the localized reconstruction methods, the interactions between the N-terminal tails (a.a. 1–20) of the fiber and PB have been elucidated (Yu et al. 2017). Form these studies, it has been suggested that the fiber N-terminal tails (FNTs) bind at the interface of

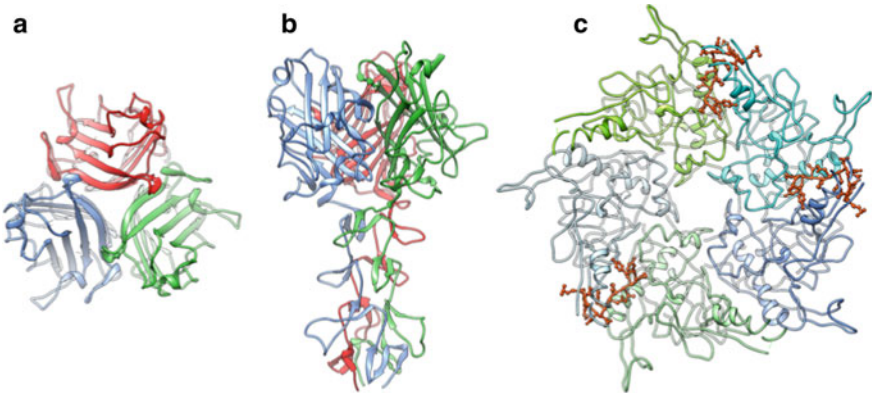


Fig. 16.4 Fiber structure and interaction of fiber N-terminal tails (FNTs) with PB pentamer. **a** Top view of the fiber showing the trimeric fiber knob of HAdV-2 (PDB:1QIU). The three independent molecules are shown in different colors **b** Side view of the fiber showing the fiber knob at the top and three β-spiral repeats of the shaft are shown at the bottom. **c** A top view of the pentameric PB showing the locations of the three FNTs of the trimeric fiber interacting with the PB (PDB:5TX1)

two adjacent PB subunits and are surface accessible (Fig. 16.4). Additionally, the FNTs exhibit elbow shaped structure(s) that interact more closely with one of the two subunits and latch onto the base of a PB structure (aa 324–330), which comprises the RGD loop. There is a consensus notion that 2 of the 3 FNTs of trimeric fiber occupy the two adjacent binding sites on PB, while the 3rd FNT binds at a site that is one site removed from either of the two sites occupied (Fig. 16.4c) (Zubieta et al. 2005; Liu et al. 2011). This arrangement of the FNTs of the trimeric fiber was compared to a "3-pronged claw" of the AdV fiber that latches on to the surface of the PB, thereby providing greater stability to the PB-fiber complex. Moreover, a 3.4 $Å^2$ per a.a. BSA estimation for the PB-FNT interaction (1821.5 $Å^2$/[520 + 14]) in HAdV-26 structure suggest the strongest inter-capsomer interactions seen in the HAdV capsid.

Structures and Interactions of Minor (Cement) Proteins

Protein IX

The only minor protein located on the exterior of AdVs is protein IX (14.4 kDa), which is present in 240 copies and forms a continuous hexagonal polypeptide network that interlaces the hexons in the entire HAdV capsid. The interactions involving IX and those with the other minor proteins located on the capsid interior stabilize the AdV capsid (see below). The protein IX molecule displays an extended structure that can be divided into three regions; N-terminal triskelion forming region (aa 1–61), linker region (aa 62–94) and C-terminal coiled-coil forming region (aa 95–134)

(Fig. 16.5a). The residue numbers indicated in the figure correspond to protein IX of HAdV-26 that is a few (six) residues shorter than its counterpart in HAdV-5. In the context of the assembled adenovirus capsid, protein IX molecules form two signature oligomeric substructures: triskelions formed by the N-terminal region, while the C-terminal coiled-coils form a 4-helix bundle (4-HLXB) (Fig. 16.5b, c). The triskelions, located at the 3-fold junctions (Fig. 16.5d, light-blue triangles) within the group of nine hexons (GON) substructure of the icosahedral facets, display perfect 3-fold symmetry and strengthen hexon-hexon interactions (Reddy 2017; Matteson et al. 2018). While the 4-HLXB structures are located near the edges of the facets and mediate interactions between the GON substructures (Fig. 16.5d, light-blue rectangles). The remarkable characteristic of the 4-HLXB is that it is made of homopolypeptides arranged 3 in parallel (Fig. 16.5b, c; shown in purple, cyan, skyblue) and 1 in anti-parallel orientation (Fig. 16.5b, c; shown in darkblue). Moreover, each of

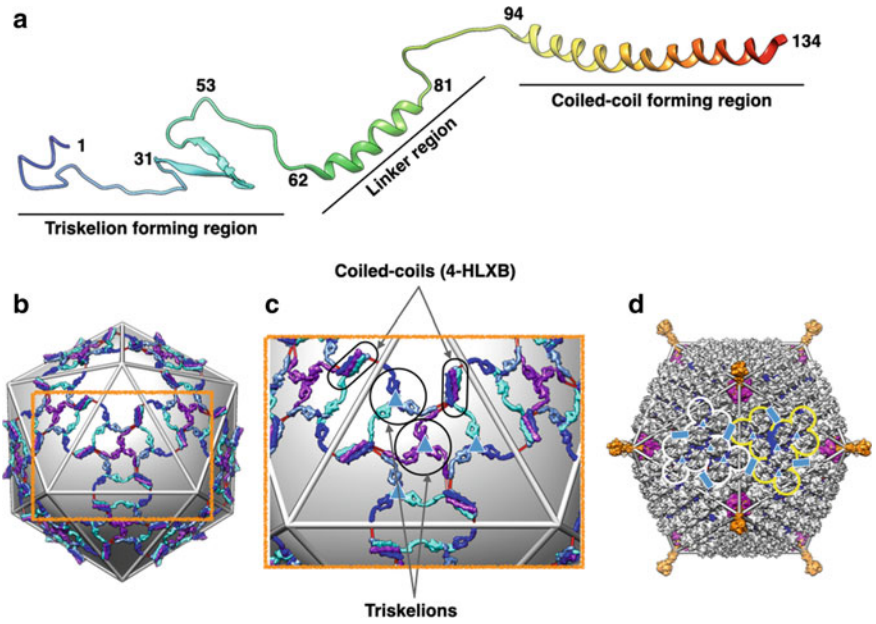


Fig. 16.5 Structure and polypeptide network of IX molecules in human adenoviruses. **a** A ribbon diagram showing the extended structure of a protein IX molecule with different subregions for forming the characteristic oligomeric substructures identified. **b** Top view of the continuous hexagonal network comprising 240 IX molecules displayed on the surface of an icosahedron (gray). The four structurally unique IX molecules are shown in different colors. **c** A zoomed-in view of the orange rectangular region in the panel B identifying the locations of the N-terminal triskelion and the C-terminal coiled-coil (4-HLXB) regions. **d** The locations of the triskelion and 4-HLXB structures, in the context of the entire HAdV particle, are identified by the blue triangles and rectangles respectively. The white and yellow color line tracings represent the outlines of the two adjacent group of nine hexon (GON) substructures

the helices is contributed by a protein-IX molecule that belongs to a different triskelion structure. Of note, the anti-parallel helix comes from a IX molecule that belongs to a triskelion located in the adjacent facet. This organization of the IX molecule is unique to human adenoviruses that belong to the genus Mastadenovirus of the Adenoviridae family, as opposed to animal adenoviruses that form a trimeric helical bundle positioned on top of the triskelion structures (Cheng et al. 2014; Schoehn et al. 2008; Matteson et al. 2018). It is noteworthy that the protein IX network mainly stabilize the hexons in the GON substructures with little or no contacts with the PPHs (peripentonal hexons) (Fig. 16.5b–d). Although the protein IX network interlaces hexon trimers throughout the entire HAdV capsid, it is not necessary for forming adenovirus virions. However, it has been shown that HAdVs devoid of the protein IX display lower virion stability and reduced infectivity (Colby and Shenk 1981; Campos et al. 2004).

Moreover, because of its surface accessibility, protein IX has been one of the proteins of choice for attaching and displaying foreign antigens of interest on the HAdV surface for the purposes of creating vaccines (Bayer et al. 2010; Dmitriev et al. 2002; Vellinga et al. 2004; Campos et al. 2004). A number of groups have attached different proteins and peptides to the C-terminus of IX to be displayed on the surface of HAdV virions with considerable success (Vujadinovic et al. 2018; Vujadinovic and Vellinga 2018). However, it has been suggested that the unique parallel and antiparallel packing and orientation of the C-terminal coiled-coils in the 4-HLXB may not lend themselves to be ideal for antigen display. This is because the large attachments to the C-terminus sterically interfere with the neighboring hexons and thereby disrupt the formation of 4-HLXB sub-structure and may potentially destabilize the HAdV capsid (Matteson et al. 2018). It has been hypothesized that the trimeric coiled-coil arrangement of the C-terminal protein IX helices seen in animal adenoviruses might be better suited for the antigen display (Matteson et al. 2018; Flint 2017).

Protein VIII

Among the AdV minor proteins, protein VIII (24.7 kDa) is highly conserved both in terms of sequence and structure. This protein is located on the capsid interior and present in 120 copies that undergo maturation cleavage by the adenoviral protease (AVP) at two locations (G110-G111, A157-G158), which are conserved between different HAdVs. Displaying an overall obtuse triangular structure spanning 100 Å in length and 30 Å height, each VIII molecule interacts with the bases of four hexons on the capsid interior (Fig. 16.6a). One of the striking interactions of VIII is "molecular clamping" that disrupts the formation of native β-strands in hexon subunits between residues 937–942 and 945–952 and replacing them with new β-strands involving its own residues (28–36 and 105–110) in two different hexons (Yu et al. 2017). These interactions hold the hexons in the GON structure together. Furthermore, the wide end of the obtuse triangle formed by the N and C-termini of VIII (1–9; 197–227),

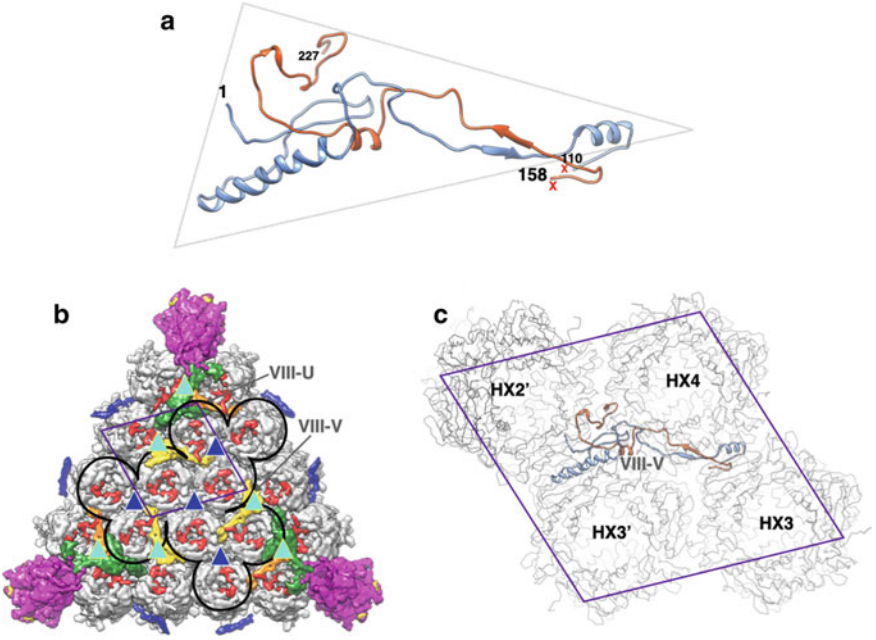


Fig. 16.6 Structure and molecular clamping interactions of VIII in human adenoviruses. **a** A ribbon diagram showing the obtuse triangular shape of protein VIII. The location of protease cleavage sites at residues 110 and 158 are identified by Xs shown in red. The two cleaved peptide fragments 1–110, 158–227 are identified by light blue and orange colors respectively. **b** Interior view of the icosahedral facet showing the locations of the structurally unique protein VIII molecules shown in orange and yellow colors. Hexons and PBs are shown in gray and magenta colors respectively, while IIIa and fragments of VI are shown in dark green and red colors, respectively. One of the protein-IX coiled-coils is shown in blue. The cyan colored triangles identify the locations of the local 3-fold axes, where VIII molecules interact with the hexons, while the dark blue triangles indicate the locations of the protein-IX triskelions on the capsid exterior. **c** A closeup view of the rhomboidal area in panel B showing a VIII (V) molecule (colored yellow) contacting the four hexons depicted in light gray backbone traces. Hexons, HX2' and HX3' belong to the adjacent GON (facet)

located at the local 3-fold axes of symmetry along the PPH-PPH and GON-GON interfaces, glue the PPHs (by VIII-U) and hexon-2 with hexon-4 from the neighboring GONs (by VIII-V) mainly through hydrophobic interactions (Fig. 16.6b, c). Thus, the protein VIII molecules play a critical role in cementing the hexons together within and between the GONs on the capsid interior. The above local 3-fold junctions located at the GON-GON interfaces are distinct from the 3-fold junctions found within the GON substructure, which are strengthened by the protein-IX triskelion structures on the capsid exterior (Figs. 16.5c and 16.6b).

Protein IIIa

Largest among the minor proteins (65 kDa), IIIa is present in 60 copies in adenovirus virions. Assigning the location of IIIa was contentious (Stewart et al. 1993; San Martin et al. 2008; Liu et al. 2011; Reddy and Nemerow 2014; Campos 2014; Condezo et al. 2015; Yu et al. 2017; Dai et al. 2017). However, thankfully, it has been resolved and now it is clear that the entire molecule of IIIa is located on the capsid interior, underneath the vertex region. The 2/3rd of IIIa (~400 aa out of 585 aa) has been resolved and displays entirely helical structure that folds into 3 domains: N-terminal domain (NTD), middle domain (MDLD) and an appendage domain (APD) (Fig. 16.7a) (Yu et al. 2017). The structure of IIIa is also well conserved between HAdV-5 and HAdV-26 structures (RMSD: 1.35 Å with sequence identity 78%). It has been suggested that the overall shape of IIIa resembles seahorse shape (Liu et al. 2010). Five molecules of IIIa are clustered underneath each vertex region in stabilizing the PPHs. Specifically, one molecule of IIIa along with a molecule of VIII glue a pair of PPHs together and joins them with the GON structure (Fig. 16.7b). In particular, a pair of 5-fold related PPHs are held together by strong interactions between the NTDs of two adjacent IIIa molecules bound to the respective PPHs

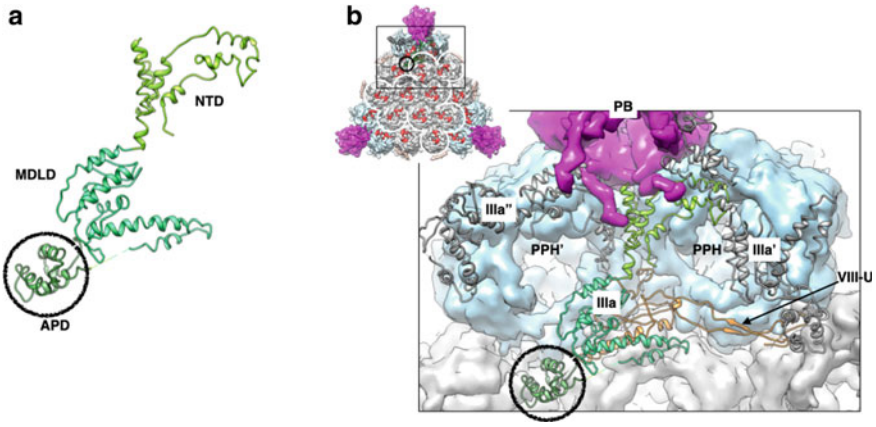


Fig. 16.7 Structure and cementing interactions of protein IIIa in human adenoviruses. **a** A ribbon diagram showing the domain organization of the ordered region (1–390) of IIIa. The three different domains N-terminal domain (NTD), middle domain (MDLD) and appendage domain (APD) are labeled. The APD is identified with a black circle. **b** A zoomed-in view of the rectangular region in the inset. Three molecules of IIIa are labeled and shown in ribbon representation (one in green and two symmetry mates in gray) interacting with the bases of the hexons, shown in surface representation, on the capsid interior. The reference IIIa molecule shown in green is overlaid on top a VIII-U molecule (orange ribbon), which together mediate the interactions between two PPHs with the hexons in two adjacent GONs. In particular, the APD domain mediates interactions at a 3-fold junction formed by a PPH and HX4 and HX2 from two neighboring GONs. Of note, the VIII-U molecules associated with the gray IIIa molecules are not shown for the purpose of clarity. Part of the PB (magenta color) is shown interacting with the NTD of IIIa

(Fig. 16.7b). Residues 30–60 of one IIIa interact closely with the residues 82–140 of the symmetry related molecule. Five sets of these interactions hold the 5 PPHs together in place. These interactions along with those involving VIII are critical for cementing the PPHs underneath the vertex region. In addition, the same pair of IIIa molecules (above) also interact with the N-terminal of residues (23–37) of PB. The middle domain (MDLD) consisting of residues 134–301 primarily interacts with the VIII-U molecule by overlaying on top of the wide end of the obtuse triangle, which together mediate interactions between two PPHs at the opposite end of the interface stabilized by the two NTD domains. The newly identified helical domain (APD, encircled in Fig. 16.7a, b) composed of amino acids 314–390 is located at a distinct 3-fold junction formed by a PPH and two hexons (HX4 and HX2) from the neighboring GON facets. This unique organization of IIIa and VIII molecules underneath the vertex region suggests that VIII molecules would bind the hexons first, followed by overlaying of IIIa molecule on top of VIII molecule at the vertex region, thereby fortifying the interactions between the PPHs and GON substructures. Of note, the complex between IIIa and VIII molecules is found only underneath the icosahedral vertex region, even though another VIII is present at a different location that stabilizes the hexons across GON-GON interface.

Protein VI

Protein IV is a multifunctional molecule that is involved in a number of key events in the AdV lifecycle (Greber et al. 1993; Meier and Greber 2003; Wodrich et al. 2003; Wiethoff et al. 2005; Suomalainen and Greber 2013; Mangel and San Martin 2014; Luisoni et al. 2015; Baniecki et al. 2001). One of the important functions of VI during the biogenesis of AdV is to transport hexon subunits from cytoplasm into nucleus for the assembly and generation of new virions. It was not known until recently where exactly VI binds in hexons, which was revealed by mass-spectrometry and structural studies (Snijder et al. 2014; Reddy and Nemerow 2014). It is the processed segment of N-terminal residues 1–33 of VI that bind near the entrance of the hexon cavity, which is exposed to the capsid interior (Fig. 16.8a) (Yu et al. 2017). The primary binding site for the N-terminus (NT) of VI in the hexon cavity is located at the interface of two hexon subunits formed by aa 48–54 of one subunit and aa 30–37 of the neighboring subunit near the entrance of the hexon cavity (Fig. 16.7a, b). Therefore, there will be three equivalent binding sites available for VI at the base of the trimeric hexon. According to the location of structurally unique hexons (1–4) (Fig. 16.1a), there is density for 2–3 copies of NT of VI in the hexon cavities, which suggest that there will be at least 480 (240 × 2) VI molecules present in a HAdV virion. Based on the structure of HAdV-26, we have suggested that the very N-terminal (NT) residues of VI (aa 3–10) are buried in deep in the hexon cavity, while the VI residues aa 26–30 are exposed to outside of the hexon cavity. This is consistent with the requirement that the cleavage at 33–34 is accessible to the protease for the processing. However, a recent high-resolution structure of HAdV-5 by Hong Zhou colleagues suggest the

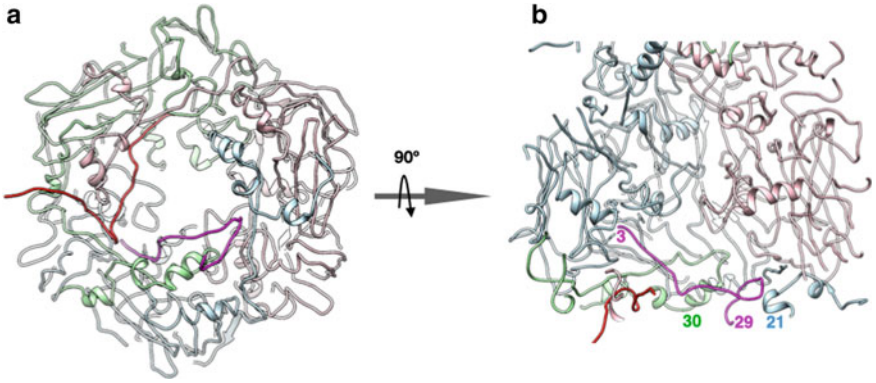


Fig. 16.8 The location and interactions of N-terminal (NT) segment of VI in the hexon cavity. **a** The location of two NT segments of VI, shown in red and magenta colors, found at the entrance of internal cavity of the peripentonal hexon (PPH). **b** A side view of the panel A showing the location of one of the NT segments reaching deep into the hexon cavity

opposite chain polarity for the VI segment for similar looking density, which places the cleavage site buried deep inside the hexon cavity (Dai et al. 2017). It is unclear how, in this case, the protease would access the cleavage site, during the maturation of AdV virions. In addition, they have identified 35 residues belonging to one of the membrane lytic fragments of mature VI (aa 109–143) in the middle of the hexon cavity. Moreover, remarkably, these authors suggested that a short fragment of VII (aa 14–24) would occupy one of the binding sites, where the NT segment of protein VI binds in the hexon cavity. Going forward, the significance of the location of the NT fragments of VI and VII in the context of adenovirus biology has to be understood.

Concluding Remarks

Unlike simple nonenveloped viral capsids that are usually made of multiple copies of a single gene product, adenovirus capsids are made of 7 different (3 major and 4 minor) capsid proteins that occur in multiple copies. Not counting the scaffolding proteins, all in all ~2800 protein molecules come together to form an adenovirus virion encapsidating the ~36 Kbp geome. How this many "moving parts" would assemble to form identical AdV virions repeatedly is a mystery to be solved. Much progress has been made in the past decade resolving the minor proteins and their network of cementing interactions. However, there a few structural details still be resolved. Some of these are: (1) how the full length membrane lytic domain of VI could be accommodated and later dislodged from the hexon cavity, (2) how the dsDNA is packaged into adenovirus virions and the details of its structure and organization and (3) how only a few copies of adenoviral protease can cleave thousands

of minor and core proteins present in the capsid interior. Future research in this area could reveal the answers to these unsolved questions.

Acknowledgements We are grateful to Glen Nemerow for the previous collaborative work on the X-ray crystallography studies on human adenovirus. This work was partially supported by the NIH grant R21 AI146644 to V.S.R and M.A.B.

References

Baniecki ML, McGrath WJ, McWhirter SM, Li C, Toledo DL, Pellicena P, Barnard DL, Thorn KS, Mangel WF (2001) Interaction of the human adenovirus proteinase with its 11-amino acid cofactor pVIc. Biochemistry 40(41):12349–12356

Bayer W, Tenbusch M, Lietz R, Johrden L, Schimmer S, Uberla K, Dittmer U, Wildner O (2010) Vaccination with an adenoviral vector that encodes and displays a retroviral antigen induces improved neutralizing antibody and CD4+ T-cell responses and confers enhanced protection. J Virol 84(4):1967–1976. https://doi.org/10.1128/JVI.01840-09

Benevento M, Di Palma S, Snijder J, Moyer CL, Reddy VS, Nemerow GR, Heck AJ (2014) Adenovirus composition, proteolysis and disassembly studied by in-depth qualitative and quantitative proteomics. J Biol Chem. https://doi.org/10.1074/jbc.M113.537498

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(14):7727–7736. https://doi.org/10.1128/JVI.78.14.7727-7736.2004

Campos SK (2014) New structural model of adenoviral cement proteins is not yet concrete. Proc Natl Acad Sci USA 111(43):E4542-4543. https://doi.org/10.1073/pnas.1415364111

Campos SK, Parrott MB, Barry MA (2004) Avidin-based targeting and purification of a protein IX-modified, metabolically biotinylated adenoviral vector. Mol Ther 9(6):942–954. https://doi.org/10.1016/j.ymthe.2004.03.006

Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks 3rd CL, Reddy VS (2009) VIPERdb2: an enhanced and web API enabled relational database for structural virology. Nucleic Acids Res 37 (Database issue):D436–442

Cheng L, Huang X, Li X, Xiong W, Sun W, Yang C, Zhang K, Wang Y, Liu H, Huang X, Ji G, Sun F, Zheng C, Zhu P (2014) Cryo-EM structures of two bovine adenovirus type 3 intermediates. Virology 450–451:174–181. https://doi.org/10.1016/j.virol.2013.12.012

Colby WW, Shenk T (1981) Adenovirus type 5 virions can be assembled in vivo in the absence of detectable polypeptide IX. J Virol 39(3):977–980

Condezo GN, Marabini R, Ayora S, Carazo JM, Alba R, Chillon M, San Martin C (2015) Structures of adenovirus incomplete particles clarify capsid architecture and show maturation changes of packaging protein L1 52/55k. J Virol 89(18):9653–9664. https://doi.org/10.1128/JVI.01453-15

Dai X, Wu L, Sun R, Zhou ZH (2017) Atomic structures of minor proteins VI and VII in human adenovirus. J Virol 91(24). https://doi.org/10.1128/JVI.00850-17

Dmitriev IP, Kashentseva EA, Curiel DT (2002) Engineering of adenovirus vectors containing heterologous peptide sequences in the C terminus of capsid protein IX. J Virol 76(14):6893–6899

Flint SJ (2017) Viral moulds and cement: how interactions among human adenovirus hexons and their protein IX cement may buttress human adenovirus particles. J Mol Biol 429(18):2752–2754. https://doi.org/10.1016/j.jmb.2017.07.008

Greber UF, Willetts M, Webster P, Helenius A (1993) Stepwise dismantling of adenovirus 2 during entry into cells. Cell 75(3):477–486. https://doi.org/10.1016/0092-8674(93)90382-z

Kundhavai Natchiar S, Venkataraman S, Mullen TM, Nemerow GR, Reddy VS (2018) Revised crystal structure of human adenovirus reveals the limits on protein IX quasi-equivalence and on

analyzing large macromolecular complexes. J Mol Biol 430(21):4132–4141. https://doi.org/10.1016/j.jmb.2018.08.011

Lenman A, Liaci AM, Liu Y, Ardahl C, Rajan A, Nilsson E, Bradford W, Kaeshammer L, Jones MS, Frangsmyr L, Feizi T, Stehle T, Arnberg N (2015) Human adenovirus 52 uses sialic acid-containing glycoproteins and the coxsackie and adenovirus receptor for binding to target cells. PLoS Pathog 11(2):e1004657. https://doi.org/10.1371/journal.ppat.1004657

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(5995):1038–1043. 329/5995/1038[pii]1126/science.1187433

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(5):764–774. https://doi.org/10.1016/j.jmb.2010.11.043

Luisoni S, Suomalainen M, Boucke K, Tanner LB, Wenk MR, Guan XL, Grzybek M, Coskun U, Greber UF (2015) Co-option of membrane wounding enables virus penetration into cells. Cell Host Microbe 18(1):75–85. https://doi.org/10.1016/j.chom.2015.06.006

Mangel WF, San Martin C (2014) Structure, function and dynamics in adenovirus maturation. Viruses 6(11):4536–4570. https://doi.org/10.3390/v6114536

Matteson NL, Barry MA, Reddy VS (2018) Structure-based assessment of protein-protein interactions and accessibility of protein IX in adenoviruses with implications for antigen display. Virology 516:102–107. https://doi.org/10.1016/j.virol.2018.01.007

Meier O, Greber UF (2003) Adenovirus endocytosis. J Gene Med 5(6):451–462. https://doi.org/10.1002/jgm.409

Reddy VS (2017) The role of hexon protein as a molecular mold in patterning the protein IX organization in human adenoviruses. J Mol Biol 429(18):2747–2751. https://doi.org/10.1016/j.jmb.2017.06.025

Reddy VS, Natchiar SK, Stewart PL, Nemerow GR (2010) Crystal structure of human adenovirus at 3.5 A resolution. Science 329(5995):1071–1075. 329/5995/1071[pii]1126/science.1187292

Reddy VS, Nemerow GR (2014) Structures and organization of adenovirus cement proteins provide insights into the role of capsid maturation in virus entry and infection. Proc Natl Acad Sci USA 111(32):11715–11720. https://doi.org/10.1073/pnas.1408462111

Rux JJ, Burnett RM (2000) Type-specific epitope locations revealed by X-ray crystallographic study of adenovirus type 5 hexon. Mol Ther 1(1):18–30

Rux JJ, Burnett RM (2007) Large-scale purification and crystallization of adenovirus hexon. Methods Mol Med 131:231–250. https://doi.org/10.1007/978-1-59745-277-9_17

San Martin C (2012) Latest insights on adenovirus structure and assembly. Viruses 4(5):847–877. https://doi.org/10.3390/v4050847

San Martin C, Glasgow JN, Borovjagin A, Beatty MS, Kashentseva EA, Curiel DT, Marabini R, Dmitriev IP (2008) Localization of the N-terminus of minor coat protein IIIa in the adenovirus capsid. J Mol Biol 383(4):923–934. https://doi.org/10.1016/j.jmb.2008.08.054

Schoehn G, El Bakkouri M, Fabry CM, Billet O, Estrozi LF, Le L, Curiel DT, Kajava AV, Ruigrok RW, Kremer EJ (2008) Three-dimensional structure of canine adenovirus serotype 2 capsid. J Virol 82(7):3192–3203. https://doi.org/10.1128/JVI.02393-07

Snijder J, Benevento M, Moyer CL, Reddy V, Nemerow GR, Heck AJ (2014) The cleaved N-terminus of pVI binds peripentonal hexons in mature adenovirus. J Mol Biol. https://doi.org/10.1016/j.jmb.2014.02.022

Stewart PL, Burnett RM, Cyrklaff M, Fuller SD (1991) Image reconstruction reveals the complex molecular organization of adenovirus. Cell 67(1):145–154. https://doi.org/10.1016/0092-8674(91)90578-m

Stewart PL, Fuller SD, Burnett RM (1993) Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. EMBO J 12(7):2589–2599

Suomalainen M, Greber UF (2013) Uncoating of non-enveloped viruses. Curr Opin Virol 3(1):27–33. https://doi.org/10.1016/j.coviro.2012.12.004

van Oostrum J, Burnett RM (1985) Molecular composition of the adenovirus type 2 virion. J Virol 56(2):439–448

van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature 401(6756):935–938

Vellinga J, Rabelink MJ, Cramer SJ, van den Wollenberg DJ, Van der Meulen H, Leppard KN, Fallaux FJ, Hoeben RC (2004) Spacers increase the accessibility of peptide ligands linked to the carboxyl terminus of adenovirus minor capsid protein IX. J Virol 78(7):3470–3479

Vujadinovic M, Khan S, Oosterhuis K, Uil TG, Wunderlich K, Damman S, Boedhoe S, Verwilligen A, Knibbe J, Serroyen J, Schuitemaker H, Zahn R, Scheper G, Custers J, Vellinga J (2018) Adenovirus based HPV L2 vaccine induces broad cross-reactive humoral immune responses. Vaccine 36(30):4462–4470. https://doi.org/10.1016/j.vaccine.2018.06.024

Vujadinovic M, Vellinga J (2018) Progress in adenoviral capsid-display vaccines. Biomedicines 6(3). https://doi.org/10.3390/biomedicines6030081

Wiethoff CM, Wodrich H, Gerace L, Nemerow GR (2005) Adenovirus protein VI mediates membrane disruption following capsid disassembly. J Virol 79(4):1992–2000

Wodrich H, Guan T, Cingolani G, Von Seggern D, Nemerow G, Gerace L (2003) Switch from capsid protein import to adenovirus assembly by cleavage of nuclear transport signals. EMBO J 22(23):6245–6255

Xia D, Henry LJ, Gerard RD, Deisenhofer J (1994) Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution. Structure 2(12):1259–1270. https://doi.org/10.1016/s0969-2126(94)00126-x

Yu X, Veesler D, Campbell MG, Barry ME, Asturias FJ, Barry MA, Reddy VS (2017) Cryo-EM structure of human adenovirus D26 reveals the conservation of structural organization among human adenoviruses. Sci Adv 3(5):e1602670. https://doi.org/10.1126/sciadv.1602670

Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. Mol Cell 17(1):121–135

Chapter 17
A Structural Perspective on Gene Repression by Polycomb Repressive Complex 2

Xin Liu

Abstract Polycomb Repressive Complex 2 (PRC2) is a major repressive chromatin complex formed by the Polycomb Group (PcG) proteins. PRC2 mediates trimethylation of histone H3 lysine 27 (H3K27me3), a hallmark of gene silencing. PRC2 is a key regulator of development, impacting many fundamental biological processes, like stem cell differentiation in mammals and vernalization in plants. Misregulation of PRC2 function is linked to a variety of human cancers and developmental disorders. In correlation with its diverse roles in development, PRC2 displays a high degree of compositional complexity and plasticity. Structural biology research over the past decade has shed light on the molecular mechanisms of the assembly, catalysis, allosteric activation, autoinhibition, chemical inhibition, dimerization and chromatin targeting of various developmentally regulated PRC2 complexes. In addition to these aspects, structure-function analysis is also discussed in connection with disease data in this chapter.

Keywords Polycomb repressive complex 2 (PRC2) · Chromatin · Epigenetics · Histone methylation · Gene repression · Structural biology

Introduction

Polycomb (*PC*), the founding member of the family of Polycomb Group (PcG) genes, was originally identified in *Drosophila melanogaster* to control segmentation by repressing homeotic (*HOX*) genes (Lewis 1978), which are known to dictate the development of anatomical structures in different species, like insects, mammals and plants. Compared to their counterparts in *Drosophila*, PcG proteins are conserved, but more diverse in mammals and plants (Hennig and Derkacheva 2009; Schuettengruber et al. 2017). PcG proteins impact a variety of fundamental biological processes, ranging from stem cell differentiation and X-chromosome inactivation in mammals

X. Liu (✉)
Cecil H. and Ida Green Center for Reproductive Biology Sciences, UT Southwestern Medical Center, Dallas, TX 75390, USA
e-mail: xin.liu@utsouthwestern.edu

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_17

to vernalization in plants (Wutz 2011; Whittaker and Dean 2017; Aloia et al. 2013). On the molecular level, PcG proteins assemble into two major repressive chromatin complexes, Polycomb Repressive Complex 1 (PRC1) and PRC2 (Fig. 17.1). Both PRC1 and PRC2 are histone modifying enzyme complexes; whereas PRC1 is responsible for monoubiquitination of histone H2A lysine 119 (H2AK119ub) (Wang et al. 2004a), PRC2 mediates trimethylation of histone H3 lysine 27 (H3K27me3) (Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002) (Fig. 17.1). Both H2AK119ub and H3K27me3 are repressive histone marks and in particular H3K27me3 is considered as a hallmark of silent chromatin. PcG proteins also form the Polycomb Repressive-Deubiquitinase complex (PR-DUB), which removes ubiquitin from H2AK119ub but, intriguingly, is also required for PcG-mediated gene silencing (Scheuermann et al. 2010).

Polycomb Response Elements (PREs) containing clusters of DNA motifs direct specific chromatin targeting of PRC1 and PRC2 in *Drosophila*, via cognate transcription factors (Ringrose and Paro 2007). In contrast, mammalian counterparts of *Drosophila* PREs appear to be lacking (Ringrose and Paro 2007). Recruitment of PRC1 and PRC2 can be interdependent. For example, in a hierarchical recruitment model, PRC2 is targeted to chromatin to install H3K27me3, which in turn recruits PRC1 through an H3K27me3 reader protein (Wang et al. 2004b). Conversely,

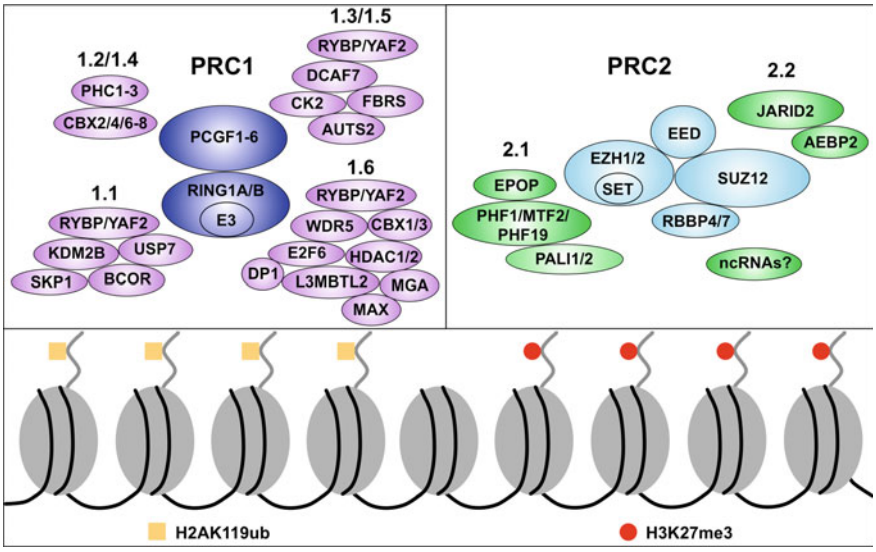


Fig. 17.1 Compositions of distinct classes of PRC1 and PRC2. Compositions of six classes of PRC1 (PRC1.1–PRC1.6) and two classes of PRC2 (PRC2.1 and PRC2.2) are shown, with the core and accessory subunits clustered separately. Histone modifications mediated by PRC1 and PRC2 are also indicated. Proteins identified from proteomics are shown as the accessory subunits of PRC1. In contrast, only proteins and ncRNAs that are known to mediate direct physical interactions with the core subunits are defined as the accessory subunits of PRC2. This more restrictive definition makes PRC2 appear to be less complex than PRC1 in the figure

H2AK119ub can be essential for the recruitment of PRC2, which contains components recognizing H2AK119ub (Blackledge et al. 2014; Cooper et al. 2014; Kalb et al. 2014; Cooper et al. 2016; Kahn et al. 2016; Blackledge et al. 2020; Tamburri et al. 2020). Hypomethylated CpG islands (CGIs) are known sites of PRC2 enrichment on chromatin (Deaton and Bird 2011). In addition, de novo recruitment of PRC2 to CGIs can be induced by inhibition of transcription (Riising et al. 2014). Indeed, CGI sequences are sufficient to recruit PRC2 in vivo in the absence of DNA methylation and transcription factor binding (Mendenhall et al. 2010; Jermann et al. 2014; Wachter et al. 2014).

PRC2 enzymatic activity is subjected to cellular regulation in both normal and diseased cells. Particularly, PRC2 binds to and is allosterically activated by H3K27me3; this positive feedback mechanism partially accounts for epigenetic inheritance, for which PRC2 association with cis-acting DNA elements is also required (Hansen et al. 2008; Margueron et al. 2009; Coleman and Struhl 2017; Laprell et al. 2017). An oncogenic histone H3K27M harboring a lysine-to-methionine mutation is linked to diffuse midline gliomas, a type of high-grade lethal brain cancer found in both children and adults; PRC2 enzymatic activity is inhibited by H3K27M, leading to a global reduction of H3K27me3 in patient samples (Chan et al. 2013; Lewis et al. 2013). Intriguingly, both gain-of-function and loss-of-function mutations of PRC2 components have been associated with human cancers (Laugesen et al. 2016). Additionally, mutations in PRC2 components also cause overgrowth disorders manifesting a variable degree of intellectual disability, including Weaver syndrome and Cohen–Gibson syndrome (Cyrus et al. 2019).

This chapter aims to provide a structural perspective on the mechanism of gene repression by PRC2, with a focus on the regulation of the enzymatic activity and chromatin recruitment of mammalian PRC2. Interested readers are also referred to many other reviews that discuss PRC2 structure and function (Simon and Kingston 2009; Margueron and Reinberg 2011; Schwartz and Pirrotta 2013; Aranda et al. 2015; Vizan et al. 2015; Chittock et al. 2017; Schuettengruber et al. 2017; Laugesen et al. 2019; van Mierlo et al. 2019; Chammas et al. 2020). Understanding composition of functional PRC2 represents one of the very first steps towards structural characterization. Below, knowledge concerning discovery, sequence feature, domain structure, complex formation and cellular function of PRC2 components is discussed.

Composition of Functional PRC2

Subunit composition and interaction of macromolecular complexes are a determinant of biological function. Multifaceted roles of PRC2 in cell development are largely correlated with the tremendous complexity and plasticity of PRC2 composition. Core subunits of PRC2 remain relatively constant, and they assemble into diverse holo complexes together with a wide spectrum of developmentally regulated accessory subunits (Fig. 17.1). Historically, components of mammalian PRC2 holo complexes were mostly identified based on homology to *Drosophila* PcG proteins or by mass

spectrometry following co-immunoprecipitation or affinity purification experiments. The accessory subunits were originally shown to be "substoichiometric", sometimes suggesting weak or transient binding. However, biochemical and structural data based on fully reconstituted PRC2 holo complexes indicated otherwise: many of the accessory subunits are stable stoichiometric components of the holo complexes. Nuclear extracts of HeLa human cervical cancer cells, HEK293/HEK293T human embryonic kidney cells and mouse embryonic stem cells (mESCs) are among the most frequently used sources for PRC2 isolation in both focused and large-scale proteomics studies (Smits et al. 2013; Alekseyenko et al. 2014; Hein et al. 2015; Huttlin et al. 2015; Maier et al. 2015; Hauri et al. 2016; Huttlin et al. 2017). In addition, mESCs and NT1/D2 human embryonal carcinoma cells offer convenient model systems for a direct comparison of the dynamic interactome of PRC2 under undifferentiated and differentiated conditions (Kloet et al. 2016; Oliviero et al. 2016). Although mammalian PRC2 has been subjected to extensive studies, the current picture of its composition may still be incomplete, for at least two reasons: cell type-specific new components of PRC2 may exist in some of the formerly unexplored cell types; furthermore, it is not impossible that certain PRC2 interactions may entail a chromatin context—not captured in chromatin-free cell extracts—and may have been missed from most of the previous proteomics studies.

RNAs and in particular a diverse collection of non-coding RNAs (ncRNAs) can be considered as non-conventional components of PRC2 holo complexes (Fig. 17.1). RNAs were shown to mediate promiscuous binding to PRC2 in cells (Khalil et al. 2009; Zhao et al. 2010; Davidovich et al. 2013; Kaneko et al. 2013), influencing critical physiological processes, such as X-chromosome inactivation by XIST (Plath et al. 2003; Silva et al. 2003; Zhao et al. 2008), repression of HOX gene clusters by HOTAIR (Rinn et al. 2007; Tsai et al. 2010), and genomic imprinting by KCNQ1OT1 (Pandey et al. 2008; Zhao et al. 2010). PRC2 preferentially binds to guanine-rich RNA motifs that may form G-quadruplex structures (Wang et al. 2017a). Canonical RNA-binding domains are not found in PRC2; instead, RNA-binding surfaces contributed by multiple PRC2 subunits—both core and accessory subunits—appear to be dispersed across the entire complex (Kaneko et al. 2014a; Long et al. 2017; Wang et al. 2017b; Zhang et al. 2019). Lacking clear binding specificity on the sequence level, whether RNAs are able to mediate specific chromatin targeting of PRC2 is still under debate (Brockdorff 2013; Davidovich and Cech 2015; Ringrose 2017). PRC2 association with RNAs is antagonistic to chromatin binding (Beltran et al. 2016; Wang et al. 2017b), and this antagonism may also underlie RNA-mediated inhibition of the enzymatic activity of PRC2 towards nucleosomal substrates (Cifuentes-Rojas et al. 2014; Kaneko et al. 2014b; Wang et al. 2017b). Ideally, structural and mutagenesis data that reveal binding specificity in three-dimensional space will help clarify a defined role of RNA molecules in PRC2 recruitment. RNA components of PRC2 will not be covered in more detail in this chapter.

The Core Complex

The PRC2 core complex from mice and humans consists of four subunits: EZH1/EZH2 (aka ENX-2/ENX-1 or KMT6B/KMT6A), EED (aka WAIT-1), SUZ12 (aka CHET9 or JJAZ1), and RBBP4/RBBP7 (aka RBAP48/RBAP46) (Fig. 17.1). EZH1 and EZH2 are paralogs and both can function as the catalytic subunit of PRC2. EZH1 and EZH2 share overall 60% sequence identity, with the SET (SU(VAR)3-9, Enhancer-of-zeste and Trithorax) domain being one of the most homologous regions (Fig. 17.2 and Table 17.1) (Abel et al. 1996; Chen et al. 1996; Hobert et al. 1996; Laible et al. 1997). Less conserved regions confer functional divergence: for example, EZH1-containing PRC2 compacts chromatin to a greater extent, whereas EZH2-containing PRC2 is better stimulated by H3K27me3 (Margueron et al. 2008; Son et al. 2013; Lee et al. 2018). EZH2 expression is correlated with active cell proliferation and EZH2 acts as the principal H3K27 methyltransferase in mESCs (Bracken et al. 2003; Lee et al. 2018). EZH1 is ubiquitously expressed and partially complements EZH2 function during mESC differentiation (Margueron et al. 2008; Shen et al. 2008). H3K27 methylation is eliminated in mESCs lacking both EZH1 and EZH2 (Shen et al. 2008; Hojfeldt et al. 2018).

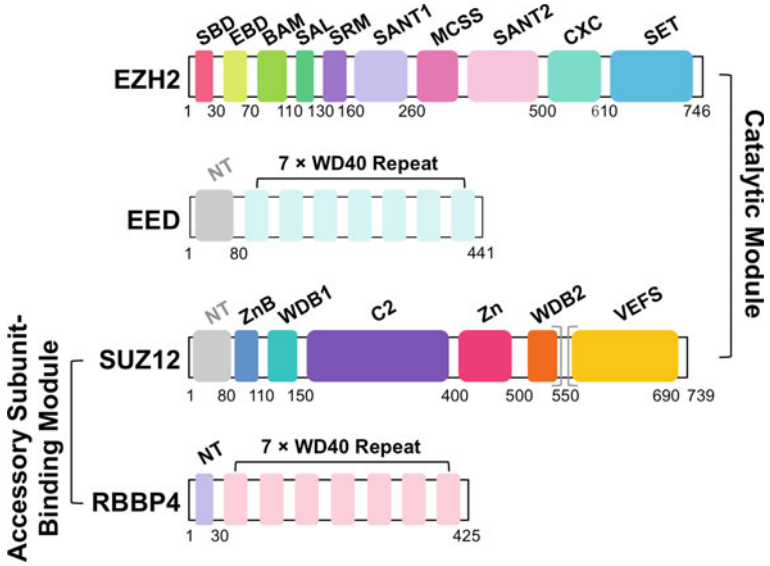


Fig. 17.2 Domain structures of the core subunits of PRC2. Domain structures of EZH2, EED, SUZ12 and RBBP4 are shown and color-coded. Domains with unknown structures are colored in grey. The subunit definition of the catalytic module and the accessory subunit-binding module is indicated. The domain structure of EZH1 is not shown, but is predicted to be similar to that of EZH2. These domain structures are defined based on sequence or structural analysis. Full names corresponding to the acronyms of protein domains are provided in Table 17.1

Table 17.1 Acronyms and full names of protein domains of the core subunits

EZH2			
SBD	SANT1-Binding Domain	EBD	EED-Binding Domain
BAM	β-Addition Motif	SAL	SET Activation Loop
SRM	Stimulation-Responsive Motif	SANT1	SWI3, ADA2, N-CoR and TFIIIB 1 (Domain)
MCSS	Motif Connecting SANT1 and SANT2	SANT2	SWI3, ADA2, N-CoR and TFIIIB 2 (Domain)
CXC	CXC (Domain)	SET	SU(VAR)3-9, Enhancer-of-zeste and Trithorax (Domain)
EED			
NT	N-Terminal (Domain)	WD40	WD40 (Domain)
SUZ12			
NT	N-Terminal (Domain)	ZnB	Zinc Finger-Binding (Helix)
WDB1	WD40-Binding (Domain) 1	C2	C2 (Domain)
Zn	Zinc Finger (Domain)	WDB2	WD40-Binding (Domain) 2
VEFS	VRN2, EMF2, FIS2, and SU(Z)12 (Domain)		
RBBP4			
NT	N-Terminal (Domain)	WD40	WD40 (Domain)

Both EED and SUZ12 are necessary for the enzymatic activity of EZH2 (Cao and Zhang 2004; Nekrasov et al. 2005). EED was found to associate with EZH1 and EZH2 in yeast two-hybrid screens (Sewalt et al. 1998; van Lohuizen et al. 1998). A large portion of EED forms a seven-bladed β-propeller WD40 domain featuring short β-sheet structural repeats of 40 amino acids (aa) ending with a tryptophan–aspartate (WD) dipeptide (Fig. 17.2 and Table 17.1) (Han et al. 2007; Stirnimann et al. 2010). EED interacts with the H3K27me3 histone mark, leading to allosteric activation of PRC2 enzymatic activity (Hansen et al. 2008; Margueron et al. 2009). SUZ12 was identified in a biochemically isolated EZH2–EED complex (Cao et al. 2002; Kuzmichev et al. 2002). Sequence analysis indicates there are a Cys_2His_2-type Zinc finger domain and a VEFS (VRN2, EMF2, FIS2, and SU(Z)12) domain in SUZ12. Structural data reveal additional functional domains in both EZH2 and SUZ12, some of which undergo dramatic conformational change in distinct states of PRC2 (*see* below) (Fig. 17.2 and Table 17.1). Like EED, RBBP4 and RBBP7 are also WD40 domain proteins (Fig. 17.2 and Table 17.1). Although frequently co-purified with the EZH2–EED complex (Cao et al. 2002; Kuzmichev et al. 2002), RBBP4 and RBBP7 are not dedicated subunits of PRC2—the two paralogs are also components of other chromatin complexes, such as Mi-2/NuRD and Sin3A histone deacetylases (HDACs) and NURF chromatin remodeling complex (Tsukiyama and Wu 1995; Zhang et al. 1997, 1998; Barak et al. 2003). RBBP4 was shown to enhance the enzymatic activity of an EZH2–EED–SUZ12 ternary complex (Cao and Zhang

2004). RBBP4 and RBBP7 are almost 90% identical in protein sequence, and it is unclear to what extent they differ from each other in PRC2 structure and function.

Isoforms of some of the core subunits exist in cells. Knowledge on isoform-specific function is limited but nonetheless worth mentioning. $EZH2^{isoform1}$ (746aa) and $EZH2^{isoform2}$ (751aa, residues 297–298 of $EZH2^{isoform1}$ replaced by another 7aa) are most frequently referred to in the literature, but functional distinction between these two isoforms has not been reported. In comparison, $EZH2^{isoform3}$ (707aa, lacking residues 83–121 of $EZH2^{isoform1}$) was found to display weaker interactions with EED and SUZ12 and to be correlated with the expression pattern of both redundant and unique gene targets in human pancreatic epithelial cells (Grzenda et al. 2013). In another study, a mouse EZH1 isoform that lacks the SET domain was shown to sequester EED in the cytosol and thereby impair the assembly of nuclear PRC2 and PRC2-dependent gene repression in differentiated C2C12 mouse myoblast cells (Bodega et al. 2017). In addition, EED isoforms with distinct N-termini that might be produced by alternative translation initiation were observed previously in HeLa nuclear extracts; however, unique PRC2 function conferred by these isoforms still remains to be elucidated (Kuzmichev et al. 2004, 2005; Martin et al. 2006).

The Holo Complexes: PRC2.1 and PRC2.2

Compositional complexity of mammalian PRC2 holo complexes is reflected by the increasing number of identified accessory subunits. These accessory subunits directly associate with the PRC2 core complex, and many of them facilitate PRC2 core complex binding to chromatin, via interactions with linker DNA, core nucleosome, or histone marks (Ballare et al. 2012; Musselman et al. 2012; Cai et al. 2013; Kalb et al. 2014; Cooper et al. 2016; Choi et al. 2017; Li et al. 2017; Wang et al. 2017b; Chen et al. 2018, 2020; Perino et al. 2018). Based on the mutually exclusive binding of the accessory subunits to the core complex, PRC2 holo complexes have so far been categorized into two classes, PRC2.1 and PRC2.2 (Fig. 17.1) (Grijzenhout et al. 2016; Hauri et al. 2016). EPOP (Elongin BC and Polycomb Repressive Complex 2-associated Protein; aka C17ORF96 or esPRC2p48), PALI1/PALI2 (PRC2 Associated LCOR Isoform 1 and 2; aka C10ORF12 as a part of PALI1) and PCL (Polycomb-Like) proteins, including PHF1 (aka PCL1), MTF2 (aka PCL2) and PHF19 (aka PCL3), are the known accessory subunits of PRC2.1 (Fig. 17.1). Those belonging to PRC2.2 include JARID2 (Jumonji/AT-Rich Interaction Domain containing 2) and AEBP2 (Adipocyte Enhancer-Binding Protein 2) (Fig. 17.1). Mutual exclusivity in PRC2 core complex binding was observed between PCL proteins and AEBP2 (Grijzenhout et al. 2016; Hauri et al. 2016; Chen et al. 2018), between PALI1 and AEBP2 (Hauri et al. 2016; Conway et al. 2018), and among EPOP, PALI1 and JARID2 (Alekseyenko et al. 2014; Beringer et al. 2016; Hauri et al. 2016; Liefke et al. 2016; Chen et al. 2018; Conway et al. 2018). PRC2.1 and PRC2.2 are complementary in maintenance of the H3K27me3 histone mark at specific chromatin loci in mESCs; combined ablation of the two classes of the accessory subunits abolishes normal

H3K27me3 pattern on chromatin, accompanied by diffuse H3K27 trimethylation likely due to mislocalization of the PRC2 core complex (Healy et al. 2019; Hojfeldt et al. 2019). In line with these observations, PCL proteins and AEBP2 were indicated to reduce mobility and prolong residence time of the respective PRC2 holo complexes on chromatin (Choi et al. 2017; Youmans et al. 2018).

PRC2.1: PHF1, MTF2 and PHF19

PHF1 and PHF19 were identified based on homology to *Drosophila* PCL protein (Coulson et al. 1998; Wang et al. 2004c). MTF2 was originally cloned as a DNA-binding protein recognizing the metal response elements (MREs) on the promoters of metallothionein (MT) genes (Inouye et al. 1994). A series of additional studies indicated that all three PCL proteins are components of PRC2 holo complexes, colocalizing with EZH2 and SUZ12 on chromatin; in many cases, knockdown of PCL proteins was shown to reduce H3K27me3 deposition on specific targets, in parallel with change of gene expression (O'Connell et al. 2001; Cao et al. 2008; Sarma et al. 2008; Walker et al. 2010; Boulay et al. 2011; Casanova et al. 2011; Li et al. 2011; Ballare et al. 2012; Brien et al. 2012; Hunkapiller et al. 2012; Cai et al. 2013). Expression of PCL proteins is developmentally regulated. For example, MTF2 is the dominant form in mESCs; in comparison, MTF2 level is downregulated during mESC differentiation to neural progenitor cells (NPCs), whereas PHF1 and PHF19 become relatively enriched (Kloet et al. 2016).

PHF1, MTF2 and PHF19 share similar domain structures: the N-terminal half of these proteins contains one Tudor domain, two PHD fingers (PHD1 and PHD2), and one EH/WH (Extended Homology/Winged-Helix) domain; located at the C-terminus is one RC/CL (Reversed Chromo/Chromo-Like) domain (Fig. 17.3 and Table 17.2) (Ballare et al. 2012; Choi et al. 2017; Li et al. 2017). The Tudor domain of all three PCL proteins binds to H3K36me3, a histone mark associated with gene body of active genes; this interaction facilitates chromatin targeting of PHF19-containing PRC2.1 for de novo silencing of active genes, a process to which H3K36me3 demethylases like NO66 and KDM2b also contribute (Abed and Jones 2012; Ballare et al. 2012; Brien et al. 2012; Musselman et al. 2012; Cai et al. 2013). The EH/WH domain of PCL proteins interacts with double-stranded linker DNA, enhancing chromatin association of PRC2.1 (Choi et al. 2017; Li et al. 2017). Furthermore, DNA helical shape was proposed to dictate specific DNA binding by MTF2 (Perino et al. 2018). The RC/CL domain is responsible for the stable binding of PCL proteins to the PRC2 core complex (Ballare et al. 2012; Choi et al. 2017; Chen et al. 2018). In the case of MTF2 and PHF19, the RC domain was also shown to stabilize a dimeric state of PRC2.1 to promote chromatin binding, possibly through an avidity effect (Chen et al. 2020).

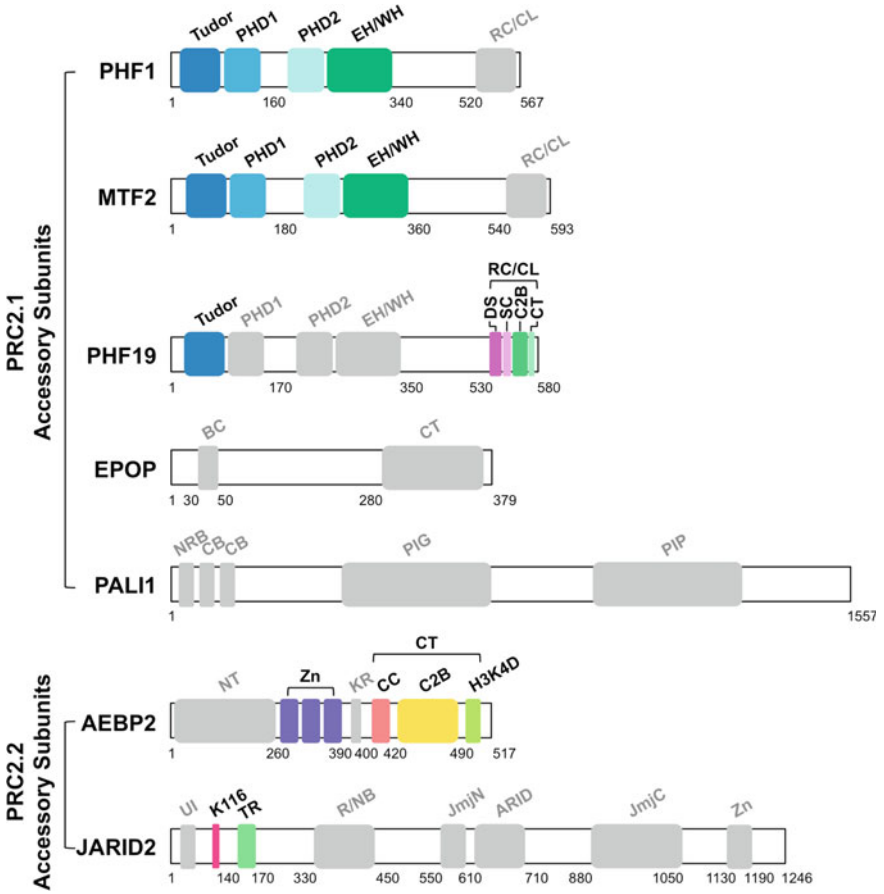


Fig. 17.3 Domain structures of the accessory subunits of PRC2. Domain structures of the accessory subunits of PRC2.1 (PHF1, MTF2, PHF19, EPOP, PALI1) and PRC2.2 (AEBP2 and JARID2) are shown and color-coded. Domains with unknown structures are colored in grey. PALI2 is less characterized and its domain structure is not included. These domain structures are defined based on sequence or structural analysis. Full names corresponding to the acronyms of protein domains are provided in Table 17.2

PRC2.1: EPOP and PALI1/PALI2

EPOP and PALI1 (the C10ORF12 part) were originally noted in a pool of proteins that associate with certain core subunits of PRC2 (Zhang et al. 2011; Smits et al. 2013; Alekseyenko et al. 2014; Maier et al. 2015; Hauri et al. 2016; Kloet et al. 2016; Oliviero et al. 2016). Proteomics on EPOP or PALI1 pulldowns indicate they are exclusive members of PRC2.1 and are incompatible between themselves in PRC2 binding (Alekseyenko et al. 2014; Beringer et al. 2016; Hauri et al. 2016;

Table 17.2 .

PHF1/MTF2/PHF19			
Tudor	Tudor (Domain)	PHD1	Plant Homeodomain (Finger) 1
PHD2	Plant Homeodomain (Finger) 2	EH/WH	Extended Homology/Winged-Helix (Domain)
RC/CL	Reversed Chromo/Chromo-Like (Domain)	DS	Dimer Stabilization (Helix)
SC	Short Connecting (Helix)	C2B	C2-Binding (Domain)
CT	C-Terminal (Tail)		
EPOP			
BC	BC (Box)	CT	C-Terminal (Domain)
PALI1			
NRB	Nuclear Receptor-Binding (Box)	CB	CTBP-Binding (Motif)
PIG	PALI Interaction with G9A (Domain)	PIP	PALI Interaction with PRC2 (Domain)
AEBP2			
NT	N-Terminal (Domain)	Zn	Zinc Finger (Domain)
KR	K (Lysine) and R (Arginine)-rich (Motif)	CC	Central Connecting (Helix)
CT	C-Terminal (Domain)	C2B	C2-Binding (Domain)
H3K4D	H3K4 Displacement (Domain)		
JARID2			
UI	Ubiquitin Interaction (Motif)	K116	K116
TR	Transrepression (Domain)	R/NB	RNA/Nucleosome-Binding (Domain)
JmjN	Jumonji N (Domain)	ARID	AT-Rich Interaction Domain
JmjC	Jumonji C (Domain)	Zn	Zinc Finger (Domain)

Liefke et al. 2016; Conway et al. 2018). EPOP interacts with PRC2 via the CT (C-Terminal) domain, which also competes with JARID2 for PRC2 binding (Fig. 17.3 and Table 17.2) (Liefke and Shi 2015; Chen et al. 2018). A BC box located in the N-terminal portion of EPOP binds to the Elongin BC heterodimer, which together with Elongin A forms the three-subunit Elongin complex, a positive regulator of RNA polymerase II transcription elongation (Fig. 17.3 and Table 17.2) (Beringer et al. 2016; Liefke et al. 2016). EPOP colocalizes with EZH2 and SUZ12 on chromatin and moderately impedes SUZ12 binding to chromatin in mESCs (Beringer et al. 2016; Liefke et al. 2016). Loss of EPOP impairs low-level permissive transcription of PRC2 target genes in mESCs and reduces proliferation of certain human cancer cell lines (Beringer et al. 2016; Liefke et al. 2016).

PALI1 and PALI2 are vertebrate-specific proteins and they are encoded by isoforms of the *LCOR* (Ligand-dependent nuclear receptor Corepressor) and

LCORL (Ligand-dependent nuclear receptor Corepressor-Like) gene loci, respectively (Conway et al. 2018). PALI1 bears one NRB (Nuclear Receptor Binding) box, two CB (CTBP Binding) motifs, one PIG (PALI Interaction with G9A) domain and one PIP (PALI Interaction with PRC2) domain (Fig. 17.3 and Table 17.2) (Conway et al. 2018). Intriguingly, the H3K9 methyltransferases G9A and GLP were previously shown to directly interact with the PRC2 core complex and control both PRC2 targeting and H3K27 trimethylation (Mozzetta et al. 2014). PALI1 promotes the methyltransferase activity of PRC2 towards oligonucleosomes in vitro (Conway et al. 2018). The antagonistic interplay between PALI1 and AEBP2 regulates the balance of PRC2.1 and PRC2.2 activities on chromatin, which influences PRC2 target gene expression during mESC differentiation (Conway et al. 2018).

PRC2.2: AEBP2

AEBP2 was originally identified as a transcriptional repressor harboring three Gli-Krüppel Cys_2His_2-type zinc fingers in mice (He et al. 1999). AEBP2 was found in PRC2 isolated from HeLa nuclear extracts and was shown to enhance PRC2 enzymatic activity towards oligonucleosomal substrates (Cao et al. 2002; Cao and Zhang 2004). There are both short and long isoforms of AEBP2, corresponding to embryonic and somatic isoforms in mice, respectively (Kim et al. 2009). AEBP2 contains multiple functional domains (Fig. 17.3 and Table 17.2), including a Zn (Zinc finger) domain that is thought to bind DNA and confer binding specificity towards methylated CpG DNAs (Kim et al. 2009; Wang et al. 2017b), a KR (K (lysine) and R (arginine)-rich) motif that enhances nucleosome binding to PRC2 (Lee et al. 2018), and a CT (C-Terminal) domain that mediates PRC2 binding and is involved in nucleosome binding as well (Cao and Zhang 2004; Chen et al. 2018). In support of PRC1-mediated PRC2 targeting (Blackledge et al. 2014; Cooper et al. 2014; Kahn et al. 2016; Blackledge et al. 2020; Tamburri et al. 2020), AEBP2 and JARID2 were shown to recognize the H2AK119ub histone mark on chromatin (Kalb et al. 2014; Cooper et al. 2016). The long somatic isoform of AEBP2 contains a unique, intrinsically disordered NT (N-Terminal) domain of over 200 amino acids marked by several glutamate/aspartate-rich, serine-rich, and glycine-rich patches, for which function remains unknown (Fig. 17.3 and Table 17.2) (Kim et al. 2009; Chen et al. 2020).

PRC2.2: JARID2

JARID2 was identified as a cDNA clone involved in development of the nervous system in mice and humans (Takeuchi et al. 1995; Berge-Lefranc et al. 1996). Although it belongs to the JmjC (Jumonji C) domain-containing histone lysine demethylase superfamily that depends on α-ketoglutarate and iron as cofactors for catalysis, JARID2 lacks the demethylation activity due to the absence of some critical catalytic residues mediating cofactor binding (Cloos et al. 2008; Landeira and

Fisher 2011). JARID2 directly associates with the PRC2 core complex, helps recruit PRC2 to chromatin, and is required for proper differentiation of mESCs (Peng et al. 2009; Shen et al. 2009; Landeira et al. 2010; Li et al. 2010; Pasini et al. 2010). The pleiotropic roles of JARID2 in embryo development are largely connected to diverse functional domains within this protein (Fig. 17.3 and Table 17.2). From the N- to C-terminus, JARID2 harbors a UI (Ubiquitin-Interacting) motif that recognizes H2AK119ub (Cooper et al. 2016), a TR (Transrepression) domain that mediates PRC2 binding and is necessary for gene repression in vivo (Kim et al. 2003; Pasini et al. 2010; Chen et al. 2018), an R/NB (RNA/Nucleosome-Binding) domain that regulates chromatin association of PRC2 (Son et al. 2013; Kaneko et al. 2014a), a JmjN (Jumonji N) domain, an ARID (AT-Rich Interaction Domain) that plays a crucial role in chromatin targeting of JARID2 likely via direct DNA binding (Kim et al. 2003; Patsialou et al. 2005; Pasini et al. 2010), a JmjC domain, and a Zn (Zinc finger) domain that may also facilitate DNA binding (Li et al. 2010). It has not been established how the characteristic JmjN and JmjC domains of JARID2 may contribute to PRC2 function in gene repression. Notably, residue K116 of JARID2 is methylated by PRC2, and trimethylated JARID2K116 (JARID2K116me3) allosterically activates PRC2 enzymatic activity on chromatin where H3K27me3 is absent, by associating with EED at a site identical to that for H3K27me3 binding (Fig. 17.3 and Table 17.2) (Sanulli et al. 2015).

Structure-Function Analysis of PRC2

Overall Structure of PRC2.2

Electron microscopy (EM) studies have provided important insights on the overall structure of a class of PRC2 holo complexes, PRC2.2. The negative-stain EM envelope at 21 Å resolution provided an initial glimpse of the structure architecture of a PRC2.2 holo complex, PRC2–AEBP2 (Ciferri et al. 2012). With support from the chemical crosslinking and internal GFP labeling data, this structure outlines an overall bipartite organization of PRC2–AEBP2 (Ciferri et al. 2012, 2015), consisting of the catalytic module and the accessory subunit-binding module (Fig. 17.4). The catalytic module contains EZH2, EED and the C-terminal VEFS domain of SUZ12 [SUZ12(VEFS)] (Figs. 17.2 and 17.4). The VEFS domain is known to mediate EZH2 binding in *Drosophila* and mammalian PRC2 (Ketel et al. 2005; Yamamoto et al. 2004). EED and the VEFS domain of SUZ12 are minimally required for the enzymatic activity of EZH2 in *Drosophila* PRC2 (Birve et al. 2001; Ketel et al. 2005). RBBP4 and the N-terminal two-thirds of SUZ12 [SUZ12(N)] together form the accessory subunit-binding module, which is defined here based on the recent finding that it is sufficient to mediate the mutually exclusive binding of most accessory subunits of PRC2.1 and PRC2.2 (Figs. 17.2, 17.3 and 17.4) (Chen et al. 2018, 2020). The accessory subunit-binding module was also formerly referred to as the

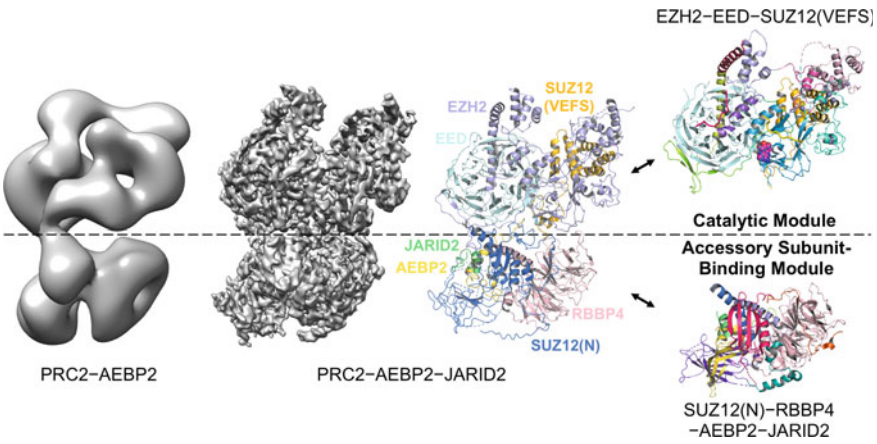


Fig. 17.4 Overall structural model of PRC2.2 from cryo-EM and crystal structures. Negative-stain EM envelope (EMD-2236) and cryo-EM map (EMD-7334) of PRC2.2 holo complexes are shown. The catalytic module and the accessory subunit-binding module are defined based on a natural demarcation of the overall structure architecture of PRC2.2. The two modules are connected through SUZ12 [*e.g.* SUZ12(N)–SUZ12(VEFS)] (PDB 6C23, holo complex cartoon on the left). Protein domains responsible for the execution and regulation of diverse PRC2 functions are resolved to atomic resolution in the crystal structures of the catalytic module and the accessory subunit-binding module (subcomplex cartoons on the right). Throughout this book chapter, PyMOL (The PyMOL Molecular Graphics System, Version 2.3.5 Schrödinger, LLC.) and Chimera (http://www.rbvi.ucsf.edu/chimera) are used for displaying structures and rendering images

nucleosome-binding module given its involvement in nucleosome binding (Nekrasov et al. 2005; Chen et al. 2018). The version of AEBP2 used in the negative stain EM structure roughly corresponds to the shorter isoform of AEBP2; from the N to C-terminus, it spans from the catalytic module to the accessory subunit-binding module (Ciferri et al. 2012).

Cryogenic-EM (Cryo-EM) maps of higher resolutions were recently generated for a modified PRC2–AEBP2 complex additionally bound to a fragment of JARID2 (residues 106–450) (Fig. 17.4) (Kasinath et al. 2018). Flexible N-terminal domains of both EED and SUZ12 as well as residues C-terminal to the VEFS domain of SUZ12 were removed to improve the resolution (Kasinath et al. 2018). An extended active state and a compact active state were resolved to an overall resolution of 3.5 and 3.9 Å, respectively, for this PRC2–AEBP2–JARID2 complex (Kasinath et al. 2018). A C-terminal portion of AEBP2 and a short JARID2 fragment are visible in these structures (Kasinath et al. 2018). Available crystal structures were docked into corresponding regions of the cryo-EM maps to facilitate model building and interpretation of unassigned electron densities (Kasinath et al. 2018). Importantly, combination of the cryo-EM maps and X-ray crystal structures provides the most accurate and complete structural information of a PRC2.2 holo complex to date (Fig. 17.4). For example, it now becomes clear how SUZ12 orchestrates the assembly of PRC2.2: the VEFS domain of SUZ12 contacts multiple domains including the SET domain

of EZH2, forming the catalytic module together with EED; the part of SUZ12 N-terminal to the VEFS domain associates with RBBP4 to form the accessory subunit-binding module, bridging AEBP2 and JARID2 to the core complex (Figs. 17.2, 17.3 and 17.4) (Justin et al. 2016; Chen et al. 2018; Kasinath et al. 2018). Structural mechanisms of PRC2 as a sum of functional domains of PRC2 components will be discussed below, when crystal structures of the catalytic module and the accessory subunit-binding module are examined in atomic details, with reference to the cryo-EM results wherever needed. In stark contrast, knowledge on the structure of PRC2.1 holo complexes has been limited, except that some latest results indicate MTF2 or PHF19-containing PRC2.1 displays a structural organization vastly different from that of PRC2.2, which may underlie a distinct mode of chromatin binding by PRC2.1 (*see* below) (Chen et al. 2020).

Structure of the Catalytic Module

Assembly

PRC2 components are not conserved in the yeast model organisms *Saccharomyces cerevisiae* (*S. cerevisiae*, baker's yeast) and *Schizosaccharomyces pombe* (*S. pombe*, fission yeast). However, PRC2 is known to mediate H3K27 methylation in other fungi, like *Neurospora crassa* (*N. crassa*, red bread mold) (Jamieson et al. 2013), *Fusarium graminearum* (*F. graminearum*, cereal pathogen) (Connolly et al. 2013), and *Cryptococcus neoformans* (*C. neoformans*, human pathogen) (Dumesic et al. 2015). The high-resolution crystal structures of an active PRC2 were first determined for the EZH2–EED–SUZ12(VEFS) catalytic module from a thermophilic fungus *Chaetomium thermophilum*, a recently sequenced model organism for structural biology (Fig. 17.5a and b) (Amlacher et al. 2011; Jiao and Liu 2015). The overall conserved *Chaetomium thermophilum* PRC2 (*ct*PRC2) is active in H3K27 trimethylation and recapitulates H3K27me3-mediated stimulation of human PRC2 enzymatic activity in a reconstituted system (Jiao and Liu 2015). Crystal structures capture *ct*PRC2 in both basal and H3K27me3-stimulated states at 2.7 and 2.3 Å resolution, respectively (Fig. 17.5a and b) (Jiao and Liu 2015). The crystal structure of the corresponding catalytic module of human PRC2 at 2.9 Å resolution was determined in the H3K27M and JARID2K116me3-bound state (Fig. 17.5c), sharing a remarkable resemblance to that of its *Chaetomium thermophilum* counterpart (Justin et al. 2016).

These *ct*PRC2 and human PRC2 structures reveal that EZH2 contains ten structurally distinct domains or motifs, which are dispersed across the entire catalytic module (Figs. 17.2 and 17.5 and Table 17.1) (Jiao and Liu 2015; Justin et al. 2016). The SBD (SANT1-Binding Domain) at the N-terminus forms an intramolecular complex with the SANT1 (SWI3, ADA2, N-CoR, and TFIIIB 1) domain (Jiao and Liu 2015; Justin et al. 2016). The SBD, EBD (EED-Binding Domain), BAM (β-Addition Motif), SAL (SET-Activation Loop), SRM (Stimulation-Responsive Motif)

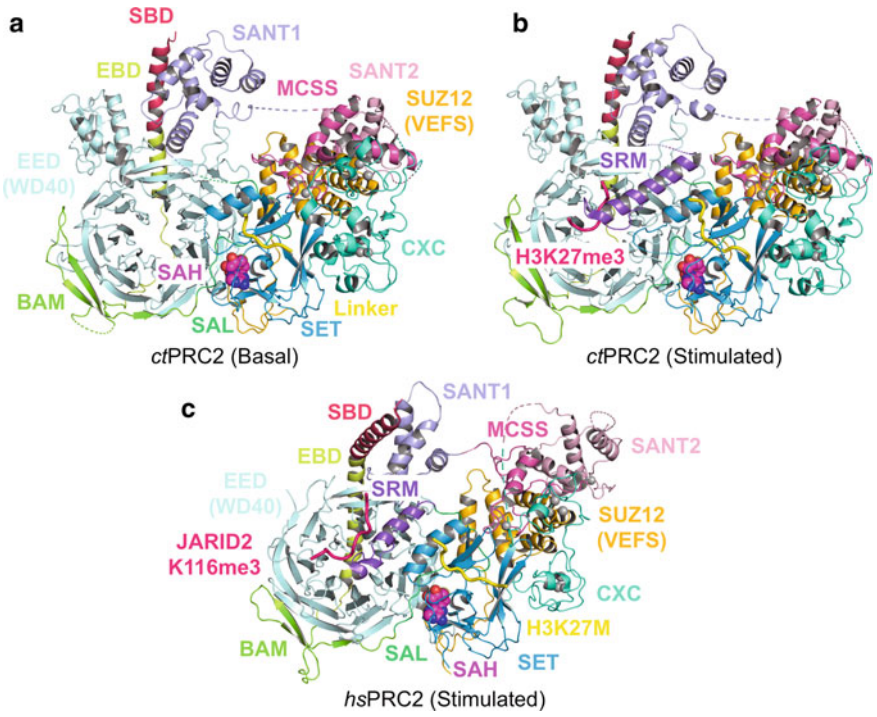


Fig. 17.5 Crystal structures of the catalytic module of *ct*PRC2 and human PRC2. Structural comparison of **a** crystal structure of *ct*PRC2 in the basal state (PDB 5KJI), **b** crystal structure of *ct*PRC2 in the H3K27me3-stimulated state (PDB 5KJH), and **c** crystal structure of human PRC2 in the JARID2K116me3-stimulated, H3K27M-inhibited state (PDB 5HYN). Corresponding domains in *ct*PRC2 and human PRC2 are colored the same, according to the color code used for the domain structures shown in Fig. 17.2. In the *ct*PRC2 structures, an internal linker sequence is trapped in the active site, mimicking the histone substrate

and SANT1 together adopt a belt-like structure that entraps EED (Jiao and Liu 2015; Justin et al. 2016). The SANT1 domain and SANT2 (SWI3, ADA2, N-CoR, and TFIIIB 2) domain are linked by the MCSS (Motif Connecting SANT1 and SANT2), which is followed by the CXC domain and the SET domain (Jiao and Liu 2015; Justin et al. 2016). The VEFS domain of SUZ12 primarily associates with the SANT2, CXC and SET domain (Figs. 17.2 and 17.5 and Table 17.1) (Jiao and Liu 2015; Justin et al. 2016). In addition to the two three-atom zinc clusters, Zn_3Cys_8His and Zn_3Cys_9, in the CXC domain, the MCSS–SANT2 region also contains a $ZnCys_3His$ and a $ZnCys_4$ zinc-binding site (Figs. 17.2 and 17.5 and Table 17.1) (Jiao and Liu 2015; Justin et al. 2016). The SAL extends to the back of the SET domain to stabilize the active conformation (Figs. 17.2 and 17.5 and Table 17.1) (Jiao and Liu 2015; Justin et al. 2016). In the presence of H3K27me3 or JARID2K116me3, the SRM helix is juxtaposed with the SET domain to bridge the stimulating signal to the active site (Figs. 17.17.2, 17.5b and c and Table 17.1) (Jiao and Liu 2015; Justin et al. 2016).

Catalytic Mechanism, Cancer Mutations and H3K27M-Mediated Inhibition of Enzymatic Activity

Two crystal structures of an isolated inactive CXC–SET region of human EZH2 at 2.0 Å resolution were available before structure determination of the active catalytic module (Antonysamy et al. 2013; Wu et al. 2013). The inactive conformation of the SET domain features an inaccessible histone substrate-binding groove blocked by the SET-I region and an incomplete SAM-binding pocket due to the largely disordered post-SET region (Fig. 17.6a) (Antonysamy et al. 2013; Wu et al. 2013). These structures provided valuable initial insights on many recurrent disease mutations, for example, mutation of residue Y641 to a phenylalanine, serine, asparagine, histidine, or cysteine residue in follicular and diffuse large B-cell lymphomas (Antonysamy et al. 2013; Wu et al. 2013; Morin et al. 2010). In addition, structure of the Zn_3Cys_8His and Zn_3Cys_9 zinc cluster-containing CXC domain was revealed for the first time (Fig. 17.6a) (Antonysamy et al. 2013; Wu et al. 2013). Similar zinc cluster structures also exist in some other lysine methyltransferases, like *N. crassa* DIM-5 and *S. pombe* CLR4 (Min et al. 2002; Zhang et al. 2002; Dillon et al. 2005). Missense mutation of the zinc-coordinating residues in the CXC domain compromises histone methyltransferase activity and is associated with myelodysplastic syndromes (MDS) (Ketel et al. 2005; Ernst et al. 2010).

Structural rearrangement of the SET domain was predicted to occur upon formation of an active PRC2 (Antonysamy et al. 2013; Wu et al. 2013). The crystal structures of the minimally active *ct*PRC2 and human PRC2 confirmed this prediction (Fig. 17.6b and c) (Jiao and Liu 2015; Justin et al. 2016). In the active conformation, the conserved NHSXXPN and EELXXDY motifs meet at the characteristic pseudoknot fold, directly contacting both SAM and the substrate lysine (Fig. 17.6b) (Dillon et al. 2005; Jiao and Liu 2015; Justin et al. 2016). Notably, the active PRC2 contains a split catalytic domain: the SAL region located in N-terminal portion of EZH2 interacts with the SET domain at the C-terminus of EZH2, and this structural configuration is likely required to maintain the active conformation of the SET domain (Fig. 17.6b) (Jiao and Liu 2015; Justin et al. 2016). Compared to the active conformation, a counterclockwise rotation of the SET-I region in the inactive conformation opens up the otherwise blocked histone substrate-binding groove (Fig. 17.6c) (Jiao and Liu 2015; Justin et al. 2016). In addition, the SAM-binding pocket becomes complete with the rotation of the SET-I and the placement of the post-SET (Fig. 17.6c) (Jiao and Liu 2015; Justin et al. 2016).

Information about substrate specificity can be partially derived from the crystal structure of JARID2K116me3 and H3K27M-bound human PRC2 (Justin et al. 2016). In addition to the specificity conferred by steric exclusion, residue R26 of histone H3 (H3R26) next to the substrate lysine is specifically recognized by residues Q648 and D652 of the SET domain (Fig. 17.6d) (Justin et al. 2016). Broader target sequences and in particular non-histone ones were recently characterized by peptide array data, suggesting a preferred substrate sequence of $(A/C/V/P)^{-3}-(A/V/L)^{-2}-(R/K)^{-1}-(K)^{0}-(F/Y/H)^{+1}-(A/V/C/T/S)^{+2}$, which is also readily rationalized based on the structure (Ardehali et al. 2017).

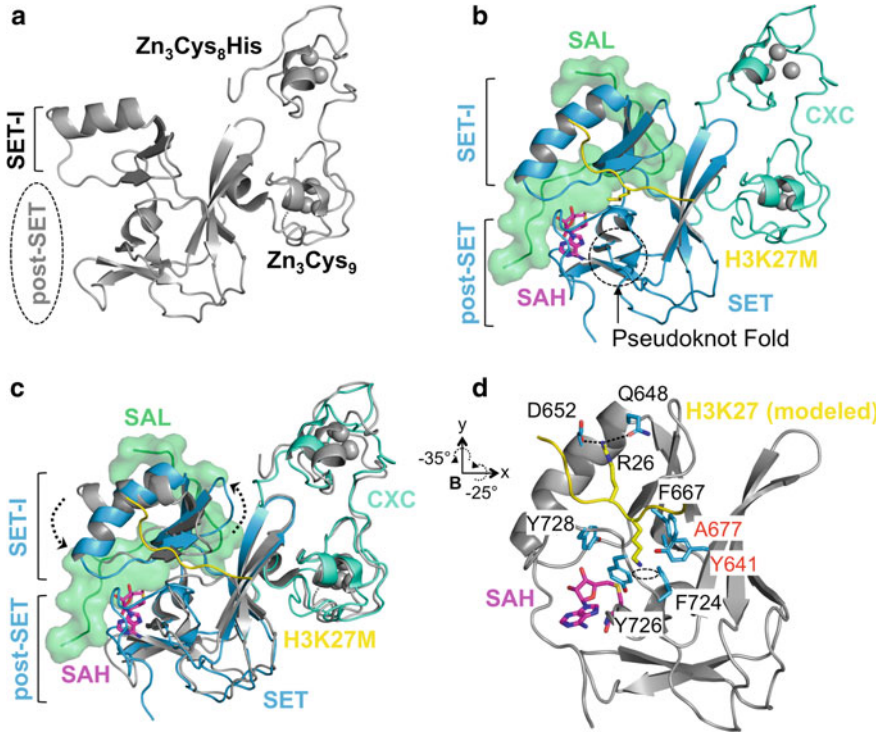


Fig. 17.6 Structural basis of the catalytic mechanism of PRC2. **a** Crystal structure of an inactive catalytic domain of EZH2 consisting of the CXC and SET domain (PDB 4MI5). Zinc clusters are labeled. The SET-I region of the SET domain is indicated. The post-SET region of the SET domain is disordered in the inactive conformation. **b** Structure of an active catalytic domain of EZH2 (PDB 5HYN). The post-SET region becomes ordered. The SAL is shown in transparent surface representation. SAH and H3K27M mimicking SAM and the histone substrate, respectively, are indicated. **c** Structure alignment of the inactive and active catalytic domain. Rotation of the SET-I region makes space for histone substrate binding and also completes the SAM-binding pocket together with the post-SET region. **d** Close-up view of the active site. The current view is rotated from the view in (**b**) according to the provided matrix. The substrate lysine is modeled *in silico* from the H3K27M structure. Aromatic residues forming the lysine-binding channel are shown in sticks representation. The dotted oval circle indicates the pore structure connecting the lysine-binding channel and SAM-binding pocket. Carbonyl oxygens and hydroxyl groups surrounding the pore structure are not shown. Sites of gain-of-function cancer mutations, Y641 and A677, are highlighted by red labels

Spatial arrangement of active site residues provides insights on the mechanism of the methyl transfer reaction, which appears to be similar to that of other lysine methyltransferases (Dillon et al. 2005; Justin et al. 2016). The model-built substrate lysine side chain is inserted through the lysine-binding channel formed mostly by aromatic residues, including Y641, F667, F724, Y726 and Y728; the nitrogen atom receiving the methyl group is positioned towards the SAM binding pocket, through a pore structure surrounded by several main chain carbonyl oxygens and side chain

hydroxyl groups of conserved tyrosine residues, which are thought to facilitate the methyl transfer process (Fig. 17.6d) (Dillon et al. 2005; Justin et al. 2016).

Mutation of some active site residues in other SET domains is known to change methylation multiplicity, by altering size and conformation of the active site (Xiao et al. 2003; Zhang et al. 2003; Collins et al. 2005). Notably, coordinated actions of the wild-type and Y641F/N/S/H/C mutant EZH2 drive H3K27 hypertrimethylation in subsets of heterozygous human B-cell lymphoma: whereas efficient in making H3K27me1 from H3K27me0, the wild-type EZH2 displays limited activity in converting H3K27me2 into H3K27me3 due to a spatially restricted active site; conversely, the Y641F/N/S/H/C mutation results in a slightly expanded active site, which disfavors H3K27 monomethylation but enhances H3K27 trimethylation (Fig. 17.6d) (Sneeringer et al. 2010). An A677G mutation proximal to residue Y641 in lymphoma cells also causes H3K27 hypertrimethylation: this mutation may promote alternative conformation of residue Y641 and enlarge the active site indirectly (Fig. 17.6d) (McCabe et al. 2012a).

The oncogenic histone mutant H3K27M was proposed to inhibit PRC2 activity, by contacting the active site directly (Chan et al. 2013; Lewis et al. 2013). Another study on PRC2 inhibition by H3K27 peptides bearing various substitutions at the K27 position suggests that the enzyme active site binds strongly to linear, hydrophobic side chains (Brown et al. 2014). Congruently, the crystal structure reveals that side chain of the methionine residue is accommodated in the lysine-binding channel by mimicking the aliphatic portion of the substrate lysine (Fig. 17.6b and d) (Justin et al. 2016). Compared to its wild-type H3K27 counterpart, an H3K27M peptide binds PRC2 with an over 10-fold higher affinity in the presence of SAM in vitro (Justin et al. 2016). In addition, H3K27me3-bound PRC2 is particularly sensitive to the inhibition by H3K27M, which also depends on SAM concentration (Stafford et al. 2018; Diehl et al. 2019). Intriguingly, a non-histone protein EZHIP (EZH1/2 Inhibitory Protein) was found to associate with PRC2 and inhibit PRC2 enzymatic activity with an H3K27M-like protein sequence in Posterior Fossa A (PFA) ependymomas (Pajtler et al. 2018; Hubner et al. 2019; Jain et al. 2019; Piunti et al. 2019).

Allosteric Activation by H3K27me3 and Weaver Syndrome Mutations

EED is a major mediator of H3K27me3-based allosteric activation of PRC2. The top surface of EED WD40 domain provides an aromatic cage, formed by residues F97, Y148, W364 and Y365, to recognize trimethylated lysine, preferentially in the A–R−K–S sequence context found for H3K27me3, H3K9me3 and H1K26me3 (Fig. 17.7a) (Margueron et al. 2009; Xu et al. 2010). The binding affinity between EED and histone peptides bearing these repressive histone marks lies in the tens of micromolar range, and aromatic cage mutations greatly diminish the binding in vitro (Margueron et al. 2009; Xu et al. 2010). The essential role of the aromatic cage in PRC2 allosteric activation by H3K27me3 in vivo is supported by the observation that mutation of the corresponding residues in *Drosophila* ESC, an equivalent of human EED, causes massive reduction of H3K27me3 level in extracts of mutant

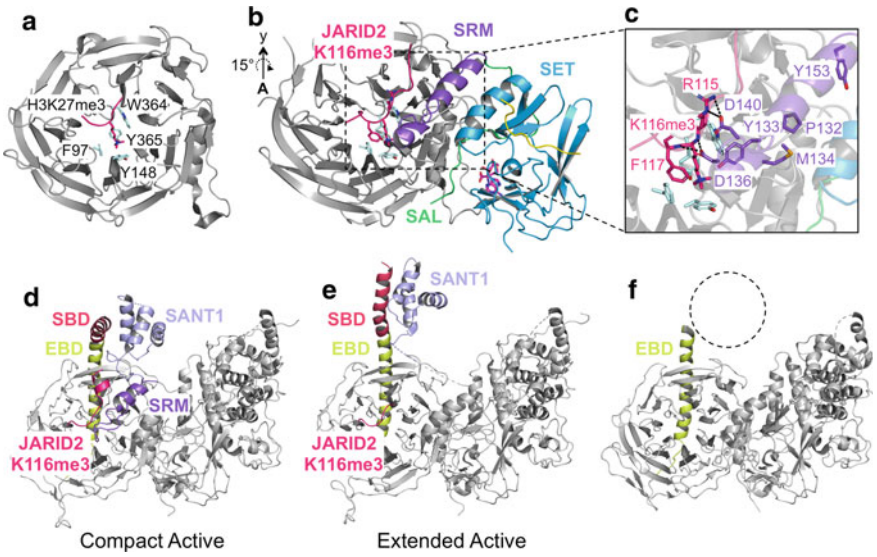


Fig. 17.7 Structural basis of the allosteric activation of PRC2. **a** Crystal structure of EED bound to an H3K27me3 peptide (PDB 3IIW). Aromatic cage residues are shown as sticks representation. **b** The SRM of EZH2. Proteins and domains from the H3K27M and JARID2K116me3-bound catalytic module of human PRC2 (PDB 5HYN) that are involved in the allosteric activation are shown. Rotation matrix that relates the current view to the view in (**a**) is indicated. **c** Close-up view of the interactions between the SRM and JARID2K116me3. Mutation of residues P132, Y133, M134 and Y153 is found in Weaver syndrome. **d** Cryo-EM Structural model of a PRC2–AEBP2–JARID2 holo complex in compact active conformation (PDB 6C23). Only the catalytic module is shown. **e** The same as (**d**) except that extended active conformation (PDB 6C24) is shown. **f** Crystal structure of the GSK126-bound catalytic module of human PRC2 (PDB 5WG6). The dotted circle indicates the disordered SBD–SANT1 region

larvae, accompanied by severe developmental defects (Margueron et al. 2009). In addition, an intact aromatic cage on EED is required for formation of H3K27me3 chromatin domains in mESCs (Oksuz et al. 2018). At a locus lacking H3K27me3, JARID2K116me3 can also stimulate PRC2 enzymatic activity for proper H3K27me3 deposition, by binding to the aromatic cage of EED with a 10-fold greater binding affinity compared with that for H3K27me3 (Fig. 17.7b and c) (Sanulli et al. 2015).

The SRM of EZH2 functions to transmit the stimulating signal from H3K27me3 or JARID2K116me3 to the SET domain (Jiao and Liu 2015; Justin et al. 2016). In the *ct*PRC2 structure, the SRM is positioned to simultaneously interact with H3K27me3 and an SET-I helix (Fig. 17.5b); the SRM becomes completely disordered in the absence of H3K27me3 (Fig. 17.5a) (Jiao and Liu 2015). Likewise, in the human PRC2 structure, the JARID2K116me3 peptide promotes the ordered conformation of the SRM, which also makes contacts with the corresponding SET-I helix (Fig. 17.7b and c) (Justin et al. 2016). The interaction between the SRM and the SET-I helix triggered by H3K27me3 or JARID2K116me3 binding may stabilize the overall conformation of the SET domain and cause dynamic changes of the active site conformation, such

that mono-, di- and trimethylation of H3K27 are all enhanced (Fig. 17.7b and c) (Jiao and Liu 2015; Justin et al. 2016). Importantly, a few de novo germline mutations of human EZH2 found in Weaver syndrome are clustered in the SRM region and are predicted to impair allosteric activation of PRC2 (Fig. 17.7c); some Weaver syndrome mutations are also mapped to the SET domain, potentially affecting the basal enzymatic activity of PRC2 (Tatton-Brown et al. 2011; Gibson et al. 2012; Al-Salem et al. 2013; Cohen et al. 2016). It remains to be studied whether these two types of Weaver syndrome mutations are associated with different disease mechanisms.

Cryo-EM maps indicate that JARID2K116me3 is bound to EED in both compact active and extended active states of a PRC2–AEBP2–JARID2 holo complex (Kasinath et al. 2018). The two states primarily differ in the position of the SBD–SANT1 intramolecular complex of EZH2 relative to EED (Fig. 17.7d and e). Compared to the extended active state, the SBD–EBD long helix is bent at the junction between the SBD and EBD in the compact active state (Fig. 17.7d and e) (Kasinath et al. 2018). This difference was also previously noted between the crystal structures of the H3K27M and JARID2K116me3-bound human PRC2 (compact active) and an inhibitor-bound human/chameleon hybrid PRC2 at 2.6 Å resolution (extended active) (Brooun et al. 2016; Justin et al. 2016). In another crystal structure of an inhibitor-bound human PRC2 at 3.9 Å resolution with a different crystal lattice, the SBD–SANT1 region becomes disordered (Fig. 17.7f), highlighting a dynamic nature of this structural unit (Bratkowski et al. 2018).

Autoinhibition and Automethylation

In addition to the basal and H3K27me3-stimuated states, the catalytic module of *ct*PRC2 also exists in an autoinhibited state in the absence of the histone substrate and SAM, which represents another distinct stable conformation (Fig. 17.8a) (Bratkowski et al. 2017). Compared to the basal and H3K27me3-stimulated states (Fig. 17.8b), the post-SET region of the SET domain becomes stably associated with the active site in the autoinhibited state, occupying the substrate-binding groove and blocking histone substrate binding (Fig. 17.8a) (Bratkowski et al. 2017). SAM binding is largely impeded in the autoinhibited state as well, due to the lack of stabilization by the post-SET region and the steric clash to the SET-I region (Fig. 17.8a and b) (Bratkowski et al. 2017). SAM binding induces local conformational change of the SET-I region and partially alleviates the autoinhibition (Bratkowski et al. 2017). A series of intramolecular interactions of EZH2 maintain the autoinhibited conformation: whereas residue F922 from the post-SET region is inserted into the lysine-binding channel, residues E840 and K852 from the SET-I region mediate hydrogen bonding interactions with the post-SET (Fig. 17.8a). Disruption of these interactions by mutagenesis results in an overactivated *ct*PRC2 (Bratkowski et al. 2017). The autoinhibition may provide a mechanism for *ct*PRC2 to sense the effective SAM or histone substrate concentration in cells (Bratkowski et al. 2017). However, critical residues mediating the autoinhibition are not conserved in human EZH2, suggesting

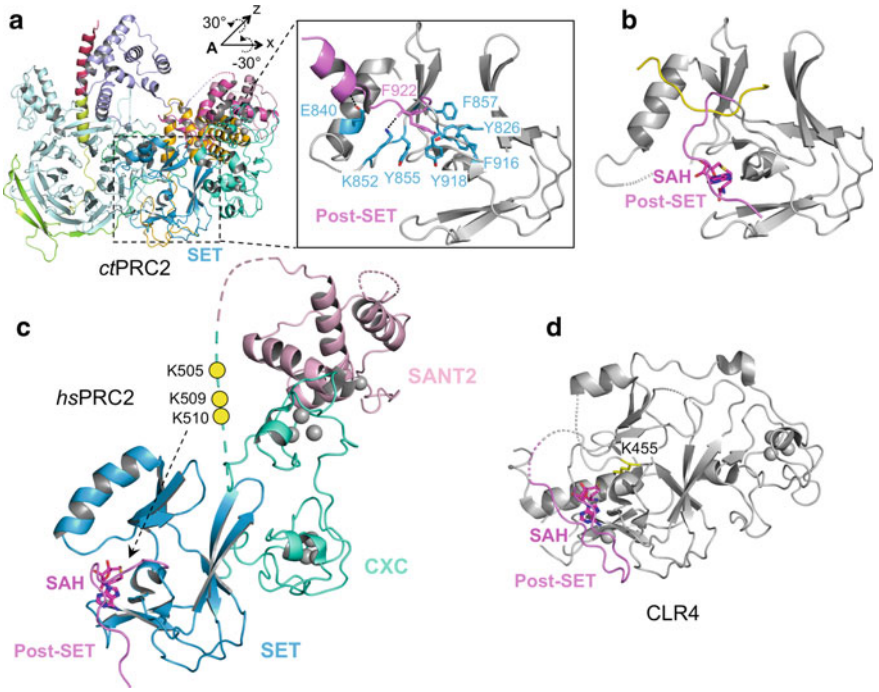


Fig. 17.8 Structural basis of PRC2 autoinhibition and automethylation. **a** Crystal structure of *ct*PRC2 in an autoinhibited conformation (PDB 5BJS). Close-up view is shown on the right, with residues that maintain the autoinhibited conformation highlighted in sticks representation. The post-SET region is colored in violet. Rotation matrix that relates the close-up view to the overall view is indicated. **b** Active conformation of the SET domain to be compared with the autoinhibited conformation. The SET domain of an active *ct*PRC2 (PDB 5KJH) is shown. The post-SET region is colored in violet. **c** Automethylation of EZH2 (based on PDB 5HYN). Three lysine residues located on a disordered loop that are subjected to automethylation are represented by filled yellow circles. The dotted arrow indicates the lysine residues are predicted to occupy the active site prior to automethylation, likely causing autoinhibition. **d** Crystal structure of the CLR4 SET domain in an autoinhibited conformation (PDB 6BOX). The view is aligned to that in (**c**). Residue K455 on a loop preceding the post-SET region occupies the active site

autoinhibition either may not exist or may involve a different mechanism for human PRC2 (Bratkowski et al. 2017).

The enzymatic activity of human PRC2 is subjected to the regulation by automethylation (Lee et al. 2019; Wang et al. 2019), reminiscent of the regulation of kinases by autophosphorylation and of histone acetyltransferases by autoacetylation (Beenstock et al. 2016; Thompson et al. 2004). PRC2 automethylation occurs intramolecularly on three lysine residues, K505, K509 and K510, on the unstructured loop connecting the SANT2 and CXC domain, (Fig. 17.8c) (Lee et al. 2019; Wang et al. 2019). Prior to the automethylation, these lysine residues may occupy the active site of EZH2, hinder histone substrate binding, and thus confer autoinhibition; the automethylation process promotes PRC2 enzymatic activity towards the histone

substrate both in vitro and in cells, conceivably by releasing the automethylation loop from the active site (Fig. 17.8c) (Lee et al. 2019; Wang et al. 2019). The proposed conformational switch during PRC2 activation by the automethylation is in line with the crystal structures of the autoinhibited and automethylated conformations of the fission yeast histone H3K9 methyltransferase CLR4 (aka SUV39H), although PRC2 enzymatic activity responds differently to alanine and arginine mutations of these automethylated lysine residues, compared to the case of CLR4 (Fig. 17.8d) (Iglesias et al. 2018; Lee et al. 2019; Wang et al. 2019).

Chemical and Stapled Peptide Inhibitors of PRC2

A family of SAM-competitive pyridone inhibitors, like GSK126, EPZ-6438 (Tazemetostat) and CPI-1205, exhibits potent (*e.g.* K_i or IC_{50} at low nanomolar range) and selective (*e.g.* > 1000-fold selective for EZH2 against many other methyltransferase families) inhibition of PRC2 enzymatic activity (Fig. 17.9a). Many of them are being tested in clinical trials for cancer treatment (McCabe et al. 2012b; Knutson et al.

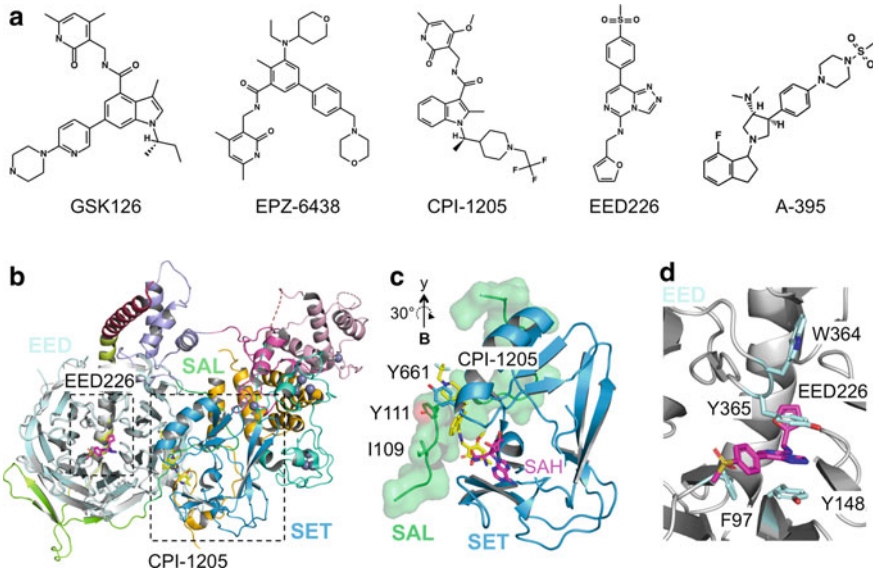


Fig. 17.9 Structural basis of chemical inhibition of PRC2. **a** Chemical structures of selected PRC2 inhibitors. **b** Aligned crystal structures of drug-bound PRC2 (PDBs 5LS6 and 5GSA). The shown structures include CPI-1205 bound to PRC2 catalytic module and EED226 bound to EED. **c** Close-up view of CPI-1205 binding pocket. Residues critical for CPI-1205 binding are shown in sticks representation and labeled. Rotation matrix that relates the current view to the view in (**b**) is indicated. **d** Close-up view of EED226 binding pocket. Aromatic cage residues mediating EED226 binding are shown. Aromatic cage residue W364 involved in H3K27me3 or JARID2K116me3 binding moves away from the EED226-binding site (compare to Fig. 17.7a)

2013; Vaswani et al. 2016). In particular, Tazemetostat (brand name Tazverik) has already been approved as a medication for metastatic or locally advanced epithelioid sarcoma and for relapsed or refractory follicular lymphoma. The unique structure of the split catalytic domain of EZH2 accounts for the remarkable selectivity of these drugs: they primarily bind to an extended pocket located on the interface between the SET and SAL domain of EZH2 that also borders EED; only a small portion of these drugs protrudes into the SAM-binding pocket and competes with SAM (Fig. 17.9b and c) (Brooun et al. 2016; Vaswani et al. 2016; Bratkowski et al. 2018). Mutation of the 'gating residues' of the drug-binding pocket, including residues I109, Y111 and Y661, results in acquired drug resistance in lymphoma cell line models (Fig. 17.9c) (Baker et al. 2015; Brooun et al. 2016; Gibaja et al. 2016).

A new class of chemical inhibitors of PRC2 became available recently. A-395 and EED226 target the aromatic cage of EED, which antagonizes H3K27me3 binding and thereby prevents the allosteric activation of PRC2 enzymatic activity (Fig. 17.9b and d) (He et al. 2017; Qi et al. 2017). These inhibitors cause similar phenotypes as the SAM-competitive inhibitors of EZH2 do in lymphoma cells; notably, they display robust PRC2 inhibition activity and strong antitumor efficacy even in the presence of the acquired mutations that confer resistance to the SAM-competitive inhibitors (He et al. 2017; Qi et al. 2017).

Stapled peptide inhibitors inhibit PRC2 enzymatic activity by disrupting complex assembly. Stapling of α-helical peptides is a strategy in drug development that allows modulation of protein-protein interactions, with the added advantage of being cell penetrating and resistant to proteolytic degradation (Verdine and Hilinski 2012). A helix of 30 amino acids of mouse EZH2 was previously mapped to be the minimal EED-binding domain (corresponding to the EBD in the catalytic module), occupying the bottom face of the EED WD40 fold in the crystal structure of a minimal EZH2–EED binary complex (Han et al. 2007). This early structural study laid the foundation for the rational design of hydrocarbon-stapled EZH2 (EBD) peptides for targeted disruption of PRC2, which impairs the growth of EZH2-dependent MLL-AF9 leukemia cells and Karpas-422 B cell lymphoma cells (Kim et al. 2013).

Structure of the Accessory Subunit-Binding Module

An Apo State

The accessory subunit-binding module of the PRC2 core complex consists of the N-terminal two thirds of SUZ12 [SUZ12(N)] and RBBP4. The *Drosophila* homologs of mammalian SUZ12 and RBBP4 are SU(Z)12 and NURF55, respectively. In the 2.3 Å resolution crystal structure of NURF55 bound to a minimal NURF55-binidng epitope of SU(Z)12, NURF55 adopts a WD40 fold and the SU(Z)12 fragment is engaged in hydrophobic interactions with a binding cleft formed by the N-Terminal (NT) helix and the PP loop of NURF55 (Fig. 17.10a) (Schmitges et al. 2011). An overlapping surface in *Drosophila* NURF55 and human RBBP7 from the Chromatin Assembly

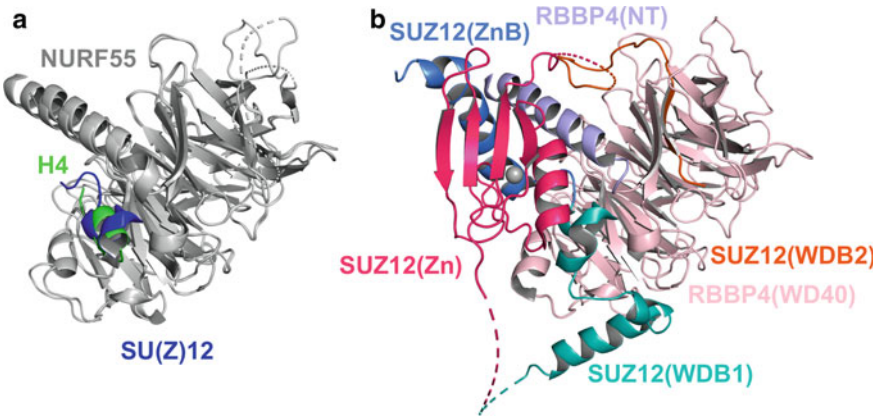


Fig. 17.10 Crystal structure of the apo accessory subunit-binding module. **a** Aligned crystal structures of *Drosophila* NURF55 bound to fragments of SU(Z)12 and histone H4 (PDBs 2YB8 and 3C9C). **b** Crystal structure of human RBBP4 bound to SUZ12(N) (PDB 5WAK). Ordered protein domains of SUZ12(N) are color coded the same as the domain structure shown in Fig. 17.2

Factor 1 (CAF-1) complex is occupied by Helix 1 of the histone fold of histone H4 based on the corresponding crystal structures of 2.4–3.2 Å resolution (Fig. 17.10a); this interaction is presumed to facilitate nucleosome assembly during DNA replication and DNA repair (Murzina et al. 2008; Song et al. 2008). A more complete structure of a human SUZ12(N)–RBBP4 binary complex at 3.3 Å resolution was solved recently: four functional domains of SUZ12 are engaged in self and RBBP4 interactions, including the ZnB (Zinc finger-Binding) helix, WDB1 (WD40-Binding domain 1), WDB2 (WD40-Binding domain 2), and Cys_2His_2-type Zn (Zinc finger) domain (Figs. 17.2 and 17.10b and Table 17.1) (Chen et al. 2018). The WDB1 harbors an equivalent of the minimal NURF55-binidng epitope of SU(Z)12 mentioned above (Fig. 17.10a and b) (Chen et al. 2018). The ZnB and Zn domains form an intramolecular complex, which also contacts RBBP4 and the WDB1 (Fig. 17.10b) (Chen et al. 2018). Notably, a large fragment of over 250 amino acids between the WDB1 and Zn domains is mostly disordered in the current structure (Figs. 17.2 and 17.10b) (Chen et al. 2018).

An AEBP2 and JARID2-Bound Monomeric State

The crystal structure of a human SUZ12(N)–RBBP4–AEBP2–JARID2 complex was determined to 3.0 Å resolution (Fig. 17.11a and b), which fits well into the cryo-EM maps of a PRC2–AEBP2–JARID2 holo complex (Fig. 17.4) (Chen et al. 2018; Kasinath et al. 2018). The crystallographic and cryo-EM data support a monomeric structure architecture of the AEBP2 and JARID2-bound accessory subunit-binding module or a corresponding PRC2.2 holo complex (Chen et al. 2018; Kasinath et al. 2018). Notably, in association with AEBP2, the large disordered region between

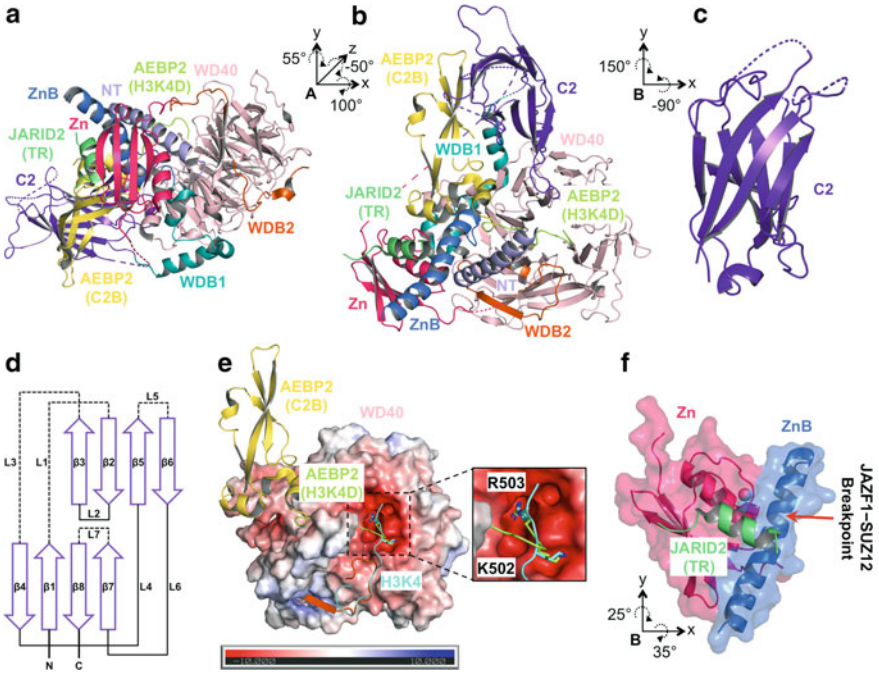


Fig. 17.11 Crystal structure of the accessory subunit-binding module of AEBP2 and JARID2-containing PRC2.2. **a** Overall view of the crystal structure of the SUZ12(N)–RBBP4–AEBP2–JARID2 heterotetrameric complex (PDB 5WAI). Protein domains are color coded as in Figs. 17.2 and 17.3. Compared to the structure of the apo complex shown in Fig. 17.10b, the C2 domain of SUZ12 becomes ordered in the current structure. **b** A different view of the heterotetrameric complex. Rotation matrix relates (**b**) to (**a**). **c** Close-up view of the C2 domain of SUZ12. The indicated rotation matrix relates (**c**) to (**b**). **d** Topology diagram of the C2 domain. **e** RBBP4 binding to the H3K4D domain of AEBP2. Electrostatic potential surface of RBBP4 is shown. Crystal structure of histone H3K4-bound *Drosophila* NURF55 (PDB 2YBA) is aligned to the current structure and the H3K4 peptide is colored in cyan. AEBP2 residues competing with H3K4 for RBBP4 binding are highlighted in the close-up view. **f** Binding surface for JARID2. Rotation matrix from (**b**) is provided. The breakpoint of SUZ12 found in the JAZF1-SUZ12 fusion oncoprotein is indicated by a red arrow

the WDB1 and Zn domains of SUZ12 becomes largely ordered (Fig. 17.11a, and compare to Fig. 17.10b), and is folded into a noncanonical type II C2 domain featuring an eight-strand β-sandwich structure, as exemplified in PLCδ, PTEN, and PI3Kα (Fig. 17.11c and d) (Chen et al. 2018). The C2 domain is known to mediate phospholipid binding and protein-protein interactions (Nalefski and Falke 1996). There are cases in which phospholipids associate with phospholipid-binding domains in chromatin complexes to impact enzymatic activity or chromatin binding activity of these complexes (Kutateladze 2012; Watson et al. 2012; Hamann and Blind 2018). However, whether the C2 domain of SUZ12 is merely a protein-protein interaction

domain or it also binds phospholipids to regulate PRC2 function remains to be further studied.

Around 100 amino acids at the C-terminus of AEBP2 (residues 407-503) are present in this heterotetrameric crystal structure, containing the C2B (C2-Binding) domain and the H3K4D (Histone H3K4 Displacement) domain (Figs. 17.3 and 17.11b and Table 17.2) (Chen et al. 2018). The C2B domain of AEBP2 forms a stable complex with the C2 domain and contacts the ZnB helix of SUZ12 as well (Fig. 17.11b) (Chen et al. 2018). The H3K4D domain sits on the top surface of RBBP4; particularly, residues K502 and R503 within this domain occupy an acidic central cavity in RBBP4 (Fig. 17.11e) (Chen et al. 2018). A comparable structural configuration of these two AEBP2 residues is also present in the cryo-EM structural model (Kasinath et al. 2018). Residues R2 and K4 on the histone H3K4 tail were previously shown to bind the corresponding acidic central cavity in *Drosophila* NURF55 in an H3K4 trimethylation-sensitive manner (Fig. 17.11e) (Schmitges et al. 2011). Structural comparison and binding assays in solution indicate AEBP2 is not compatible with histone H3K4 in RBBP4 binding (Fig. 17.11e) (Chen et al. 2018). The reason why AEBP2-containing PRC2 evolves to block H3K4 binding to RBBP4 is unclear; it is hypothesized that this structural mechanism can potentially ensure sequential recruitment of NuRD histone deacetylase and PRC2: during de novo repression of active loci, NuRD as the major binder of the unmethylated histone H3K4 tail is initially recruited to remove the acetylated histone H3K27 (H3K27ac) active histone mark, before PRC2 can be targeted to mediate histone methylation at the same site (Reynolds et al. 2012; Chen et al. 2018).

In addition to AEBP2, the crystal structure of the heterotetrameric complex also captures a JARID2 peptide (residues 147–165), corresponding to the minimal transrepression (TR) domain of JARID2 (Figs. 17.3 and 17.11f and Table 17.2) (Kim et al. 2003; Pasini et al. 2010; Chen et al. 2018). The ZnB–Zn intramolecular complex of SUZ12 is found in both the crystal and cryo-EM structural models (Chen et al. 2018; Kasinath et al. 2018). It provides a composite surface for the binding of the TR domain, which is largely folded into an α-helix structure (Fig. 17.11f) (Chen et al. 2018). The TR domain of JARID2 and the C-Terminal (CT) domain of EPOP compete for the binding of the SUZ12(N)–RBBP4 binary complex, in line with the mutually exclusive binding of JARID2 and EPOP in holo complex formation (Chen et al. 2018). In endometrial stromal tumors, a JAZF1–SUZ12 fusion oncoprotein is generated by a recurrent chromosomal translocation with a breakpoint located in the middle of the ZnB helix of SUZ12 (Fig. 17.11f) (Koontz et al. 2001; Chen et al. 2018). Based on the structural and biochemical data, JAZF1–SUZ12 is predicted to disrupt JARID2 association with PRC2, which may compromise chromatin targeting of PRC2 and contribute to the global reduction of H3K27me3 in patient samples (Ma et al. 2017; Chen et al. 2018).

A PHF19-Bound Dimeric State

In contrast to the monomeric structure of PRC2.2, data from a number of studies over the years suggest the existence of PRC2 dimers in *Drosophila*, mice and humans (O'Connell et al. 2001; Tie et al. 2003; Margueron et al. 2008; Casanova et al. 2011; Ballare et al. 2012; Son et al. 2013; Davidovich et al. 2014; Grijzenhout et al. 2016). Finally, a recent structural study provides a molecular basis for dimerization of MTF2 (based on sequence homology) or PHF19-containing PRC2.1 (Chen et al. 2020). The crystal structure of a human SUZ12(N)–RBBP4–PHF19–JARID2 complex was determined to 2.9 Å resolution, revealing the SUZ12(N)–RBBP4 accessory subunit-binding module adopts a dimeric structure architecture when bound to the RC/CL domain of PHF19 (Figs. 17.3, 17.12a and b and Table 17.2) (Chen et al. 2020). Although the TR domain of JARID2 is also present in the crystal structure, it was only used to improve the crystal quality, without interfering with PHF19 binding

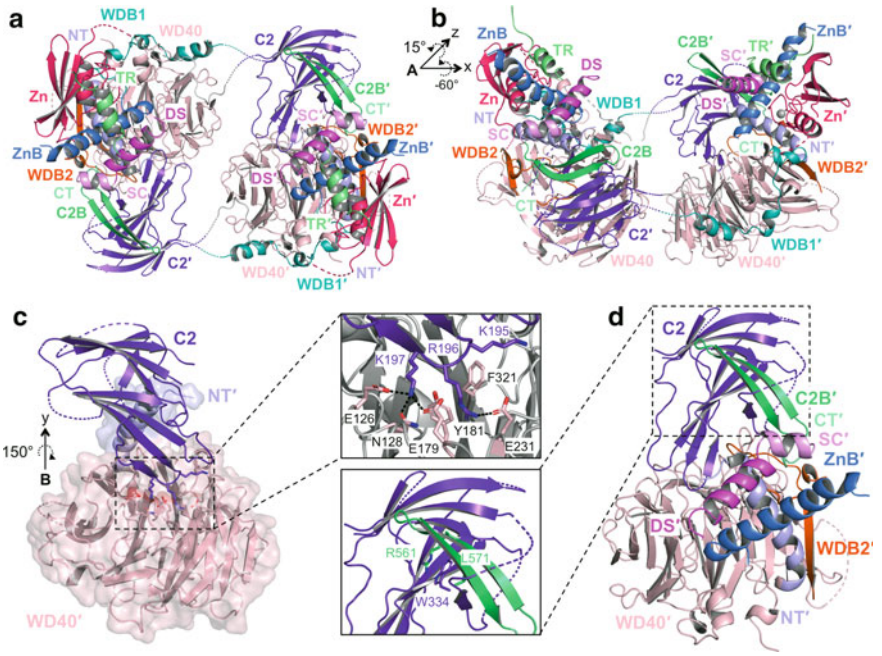


Fig. 17.12 Crystal structure of the accessory subunit-binding module of PHF19-containing PRC2.1. (**a**) Overall view of the crystal structure of the SUZ12(N)–RBBP4–PHF19–JARID2 heterotetrameric complex (PDB 6NQ3). Protein domains are color coded as in Figs. 17.2 and 17.3. Corresponding protein domains of symmetry-related protomers in the dimer are distinguished by a prime symbol in the name. **b** A different view of the heterotetrameric complex. Rotation matrix relates (**b**) to (**a**). **c** Interactions on the dimer interface. Important residues are labeled in the close-up view on the right. Rotation matrix relates (**c**) to (**b**). **d** Interactions mediating PHF19 binding to SUZ12(N)–RBBP4. The view is the same as that in (**a**). The 'R-W-L' triad is highlighted in the close-up view on the left

(Fig. 17.12a and b) (Chen et al. 2020). Intriguingly, formation of an atypical hybrid PRC2 holo complex containing both MTF2 and JARID2 was previously reported in mESCs in the absence of AEBP2 (Grijzenhout et al. 2016). More specifically, the SUZ12(N)–RBBP4 part forms a dimer by swapping the C2 domain of SUZ12 between each symmetry-related protomer (Fig. 17.12a and b) (Chen et al. 2020). On the major dimer interface, three positively charged residues, K195, R196 and K197, from the C2 domain of SUZ12 engage with the electronegative central cavity of RBBP4 through a series of hydrophobic, cation-π, charge-charge and hydrogen-bonding interactions (Fig. 17.12c) (Chen et al. 2020).

PHF19 does not participate in dimerization per se, but instead stabilizes the intrinsic dimer of SUZ12(N)–RBBP4, by associating with the C2 domain of SUZ12 from one protomer and with the ZnB and WDB2 domains of SUZ12 and the NT helix of RBBP4 from the other protomer (Fig. 17.12d) (Chen et al. 2020). The RC/CL domain of PHF19 can be further divided into four functional regions: a DS (Dimer Stabilization) helix, an SC (Short Connecting) helix, a C2B (C2-Binding) domain and a CT (C-Terminal) tail (Figs. 17.3 and 17.12d and Table 17.2). The C2B domain of PHF19 is 'locked' to the C2 domain of SUZ12 from one protomer via an 'R-W-L' triad involving residues R561 and L571 of the C2B and residue W334 of the C2; the DS and CT regions of PHF19 associate with the other protomer: the DS helix packs against the ZnB helix of SUZ12 and the NT helix of RBBP4, whereas the CT tail fits into a surface cleft formed by the WDB2 domain of SUZ12 (Fig. 17.12d) (Chen et al. 2020). Intriguingly, both MTF2 and *Drosophila* PCL contain the conserved sequence of the DS helix found in PHF19; the corresponding sequence is not conserved in PHF1, suggesting it may not stabilize the intrinsic PRC2 dimer.

Results from in vitro dimerization assays in solution are congruent with the structural observations. Mutation of residues at the dimer interface, including the K195D/R196D/K197D triple mutation and R196A single mutation of the C2 domain of SUZ12, blocks formation of the intrinsic PRC2 dimer (Chen et al. 2020). Deletion of the DS helix prevents PHF19-mediated stabilization of the intrinsic PRC2 dimer; alanine mutation of residue W334 of SUZ12 causes a similar phenotype, due to the loss of RC/CL domain binding to SUZ12(N)–RBBP4 (Chen et al. 2020). The dimer-disrupting mutations of SUZ12 also impair the binding of MTF2 or PHF19-containing PRC2 to a piece of natural CGI DNA in vitro (Chen et al. 2020). In addition, replacing the endogenous SUZ12 with the dimer-disrupting K195D/R196D/K197D triple mutant in mESCs compromises H3K27me3 enrichment on PRC2 targets, suggesting a potentially critical role of PRC2 dimerization in the chromatin recruitment of PRC2 (Chen et al. 2020).

Distinct Structure Architectures and Modes of Chromatin Binding

DNA or nucleosome-bound structures of PRC2 subcomplexes or protein domains have recently started to emerge, providing important insights on the mechanism of chromatin binding by PRC2 holo complexes. The cryo-EM structural model of an

AEBP2-containing PRC2.2 holo complex engaging an H3K27me3–H3K27 bifunctional dinucleosome at an overall resolution of 6.2 Å reveals how histone H3K27 methylation and H3K27me3-mediated allosteric activation of the enzyme may be coordinated in a chromatin context (Fig. 17.13a) (Poepsel et al. 2018). H3K27me3 is mimicked by a chemically installed methylation analog in the study (Poepsel et al. 2018). Whereas the DNA backbone of the H3K27me3 nucleosome directly contacts the SBD of EZH2, the DNA backbone of the H3K27 substrate nucleosome interacts with the CXC domain of EZH2, guiding the histone H3K27 tail to the enzyme active

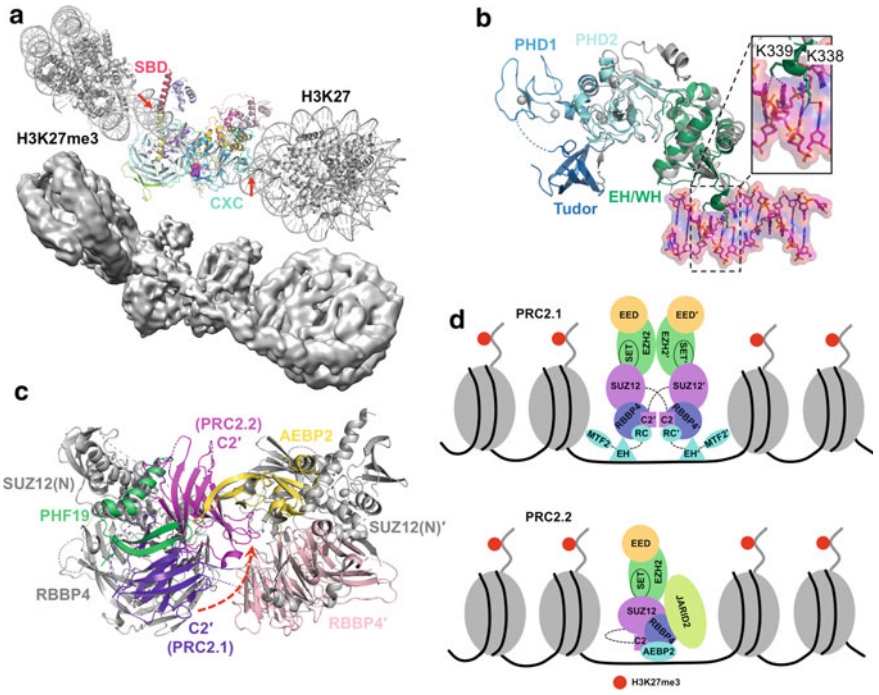


Fig. 17.13 Modes of chromatin binding of PRC2.1 and PRC2.2. **a** Cryo-EM map and structural model of a PRC2.2 holo complex, PRC2-AEBP2, bound to an H3K27me3−H3K27 bifunctional dinucleosome (EMD-7306). Only the catalytic module is resolved in the map. The SBD and CXC domains of EZH2 that contact DNA directly are indicated. **b** Crystal structure of MTF2 bound to DNA (PDB 5XFR). The Tudor, PHD1, PHD2 and EH/WH domains of MTF2 are colored according to the domain structure shown in Fig. 17.3. Crystal structure of the PHD2–EH/WH cassette of *Drosophila* PCL is aligned to the MTF2 structure and colored in gray. Interactions between the W1 loop and DNA major groove are highlighted in the close-up view. **c** Relocation of the C2 domain from the PHF19-bound to AEBP2-bound state. The two structures are aligned based on RBBP4. For clarity, only one copy of the C2 domain in the PHF19-bound dimeric structure is colored and labeled (C2′ in this case). AEBP2 binding structurally relocates the C2′ domain from the dimer interface and results in dimer disruption, as indicated by the dotted red arrow. Also refer to Fig. 17.12b. **d** Schematic of chromatin binding by the dimeric PRC2.1 and monomeric PRC2.2 holo complexes. H3K36me3 binding by PCL proteins and H2AK119ub binding by AEBP2 and JARID2 are not depicted.

site (Fig. 17.13a) (Poepsel et al. 2018). Only the catalytic module but not the accessory subunit-binding module is refined to the final resolution, together with the bound dinucleosome (Poepsel et al. 2018). 2D class averages of negative-stain EM particles indicate the AEBP2-bound accessory subunit-binding module loosely associates with a nucleosome (Poepsel et al. 2018), in line with the mononucleosome binding EMSA (Electrophoresis Mobility Shift Assay) data reported elsewhere (Chen et al. 2018). The Zn finger domain of AEBP2 is not visible in the structure, suggesting the dinucleosome used may not be optimal for the binding of this domain (Kim et al. 2009; Wang et al. 2017b). Intriguingly, it was recently proposed that the AEBP2 Zn finger might have functions other than DNA binding, based on the sequence feature and NMR solution structure (Sun et al. 2018).

The recent 2–2.5 Å resolution crystal structure of a CpG-containing duplex DNA bound to PHF1 or MTF2 harboring the Tudor–PHD1–PHD2–EH/WH cassette reveals how PCL proteins may directly facilitate chromatin targeting of PRC2.1 (Fig. 17.13b) (Li et al. 2017). Residues K323 and K324 from the W1 loop of the EH/WH domain in PHF1 are responsible for recognition of the unmethylated CpG sequence content (Li et al. 2017). The corresponding residues, K338 and K339, in MTF2 contact DNA in the same manner (Fig. 17.13b) (Li et al. 2017). Compared with the 2.4 Å resolution crystal structure of the PHD2–EH/WH cassette of *Drosophila* PCL in the absence of DNA, the partially conserved W1 loop of PHF1 and MTF2 is reshaped to insert into the major groove of the DNA, which is distorted to accommodate the W1 loop (Fig. 17.13b) (Choi et al. 2017; Li et al. 2017). DNA binding by the EH/WH domain appears to occur independently of H3K36me3 binding by the Tudor domain of PHF1 or MTF2, although both mechanisms are thought to contribute to chromatin targeting of PRC2.1 (Li et al. 2017).

Structural comparison highlights the vastly different structure architectures of the accessory subunit-binding module from PRC2.1 and PRC2.2: AEBP2 disrupts the intrinsic dimer of the accessory subunit-binding module, whereas MTF2 and PHF19 stabilizes it; in parallel, the C2 domain of SUZ12 is placed in completely different locations in the complexes to accommodate the transition between the monomeric and dimeric state (Fig. 17.13c) (Chen et al. 2020). Together with the respective accessory subunits, the accessory subunit-binding module may also control the oligomerization state of PRC2.1 and PRC2.2 holo complexes and thereby dictate the mode of chromatin binding by PRC2.1 and PRC2.2 (Fig. 17.13d) (Chen et al. 2020). As shown by the cryo-EM study, an AEBP2-containing PRC2.2 holo complex exists as a monomer (Kasinath et al. 2018). In contrast, the actual structure of PRC2.1 holo complexes is not yet available, although MTF2 or PHF19-containing ones are predicted to be a dimer based on the existing biochemical data and partial structures (Chen et al. 2020). It is proposed that PRC2 dimerization may at least enhance MTF2 or PHF19 binding to CGI DNA in pairs, possibly through an avidity effect (Fig. 17.13d) (Chen et al. 2020). Likewise, H3K27me3 or substrate nucleosome binding by the catalytic module may also be promoted in this dimeric structural framework of PRC2.1 holo complexes.

Concluding Remarks and Future Perspectives

The past decade has witnessed a growing interest and rapid progress in structural biology towards understanding the molecular basis of the assembly, catalysis and targeting of PRC2, a key mediator of gene repression during development. A series of structural studies have generated important mechanistic insights, which verify hypothetical models from functional studies, deepen the understanding of disease mechanisms, and also lay a foundation for new lines of research.

Structural data need to be interpreted critically. Whereas low-resolution structures can be valuable in offering information about the overall shape and assembly of the target, a medium to high-resolution structure is ideal to provide accurate side chain positions, for example, to allow hypotheses to be tested by specific mutations. In the 3–4 Å medium resolution range, de novo model building of large chromatin complexes like PRC2 can be found challenging and sometimes intimidating in realms of both X-ray crystallography and cryo-EM. Whenever necessary, the register of structural models should be validated and other biochemical or functional data should be taken into consideration to support model building.

Many pressing questions remain to be answered regarding structure and function of PRC2. For example, structures of nucleosome-bound PRC2.1 holo complexes are still lacking. How do functional nucleosomes communicate with each other and with the active site of PRC2.1 to influence de novo installation and spreading of the H3K27me3 repressive histone mark? Can different modes of chromatin binding by PRC2.1 and PRC2.2 be translated into distinct mechanisms of gene repression? How may ncRNAs fit into the picture of chromatin-bound PRC2.1 or PRC2.2? Another important topic is concerned with epigenetic inheritance. Does PRC2 dimerization play a role in the faithful transmission of the H3K27me3 histone mark during mitosis? Are there functional crosstalk and physical contact between PRC2 and DNA replication machinery? In answering these questions and many others, a whole new world of structural biology research on PRC2 is anticipated to unfold.

Acknowledgements I thank members of the Liu laboratory for critical reading of this book chapter. This work was supported by Welch Foundation research grant I-1790 and NIH grants GM121662 and GM136308. X.L. is a W. W. Caruth, Jr. Scholar in Biomedical Research at UT Southwestern Medical Center.

References

Abed JA, Jones RS (2012) H3K36me3 key to Polycomb-mediated gene silencing in lineage specification. Nat Struct Mol Biol 19(12):1214–1215. https://doi.org/10.1038/nsmb.2458

Abel KJ, Brody LC, Valdes JM, Erdos MR, McKinley DR, Castilla LH, Merajver SD, Couch FJ, Friedman LS, Ostermeyer EA, Lynch ED, King MC, Welcsh PL, Osborne-Lawrence S, Spillman M, Bowcock AM, Collins FS, Weber BL (1996) Characterization of EZH1, a human homolog of Drosophila Enhancer of zeste near BRCA1. Genomics 37(2):161–171. https://doi.org/10.1006/geno.1996.0537

Alekseyenko AA, Gorchakov AA, Kharchenko PV, Kuroda MI (2014) Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. Proc Natl Acad Sci USA 111(7):2488–2493. https://doi.org/10.1073/pnas.1400648111

Al-Salem A, Alshammari MJ, Hassan H, Alazami AM, Alkuraya FS (2013) Weaver syndrome and defective cortical development: a rare association. Am J Med Genet A 161A(1):225–227. https://doi.org/10.1002/ajmg.a.35660

Aloia L, Di Stefano B, Di Croce L (2013) Polycomb complexes in stem cells and embryonic development. Development 140(12):2525–2534. https://doi.org/10.1242/dev.091553

Amlacher S, Sarges P, Flemming D, van Noort V, Kunze R, Devos DP, Arumugam M, Bork P, Hurt E (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. Cell 146(2):277–289. https://doi.org/10.1016/j.cell.2011.06.039

Antonysamy S, Condon B, Druzina Z, Bonanno JB, Gheyi T, Zhang F, MacEwan I, Zhang A, Ashok S, Rodgers L, Russell M, Gately Luz J (2013) Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. PLoS ONE 8(12):e84147. https://doi.org/10.1371/journal.pone.0084147

Aranda S, Mas G, Di Croce L (2015) Regulation of gene transcription by Polycomb proteins. Science advances 1(11):e1500737. https://doi.org/10.1126/sciadv.1500737

Ardehali MB, Anselmo A, Cochrane JC, Kundu S, Sadreyev RI, Kingston RE (2017) Polycomb repressive complex 2 Methylates Elongin A to regulate transcription. Mol cell 68(5):872–884 e876. https://doi.org/10.1016/j.molcel.2017.10.025

Baker T, Nerle S, Pritchard J, Zhao B, Rivera VM, Garner A, Gonzalvez F (2015) Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. Oncotarget 6(32):32646–32655. https://doi.org/10.18632/oncotarget.5066

Ballare C, Lange M, Lapinaite A, Martin GM, Morey L, Pascual G, Liefke R, Simon B, Shi Y, Gozani O, Carlomagno T, Benitah SA, Di Croce L (2012) Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat Struct Mol Biol 19(12):1257–1265. https://doi.org/10.1038/nsmb.2434

Barak O, Lazzaro MA, Lane WS, Speicher DW, Picketts DJ, Shiekhattar R (2003) Isolation of human NURF: a regulator of Engrailed gene expression. The EMBO journal 22(22):6089–6100. https://doi.org/10.1093/emboj/cdg582

Beenstock J, Mooshayef N, Engelberg D (2016) How do protein kinases take a selfie (Autophosphorylate)? Trends Biochem Sci 41(11):938–953. https://doi.org/10.1016/j.tibs.2016.08.006

Beltran M, Yates CM, Skalska L, Dawson M, Reis FP, Viiri K, Fisher CL, Sibley CR, Foster BM, Bartke T, Ule J, Jenner RG (2016) The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res 26(7):896–907. https://doi.org/10.1101/gr.197632.115

Berge-Lefranc JL, Jay P, Massacrier A, Cau P, Mattei MG, Bauer S, Marsollier C, Berta P, Fontes M (1996) Characterization of the human jumonji gene. Hum Mol Genet 5(10):1637–1641. https://doi.org/10.1093/hmg/5.10.1637

Beringer M, Pisano P, Di Carlo V, Blanco E, Chammas P, Vizan P, Gutierrez A, Aranda S, Payer B, Wierer M, Di Croce L (2016) EPOP functionally links Elongin and Polycomb in pluripotent stem cells. Mol Cell 64(4):645–658. https://doi.org/10.1016/j.molcel.2016.10.018

Birve A, Sengupta AK, Beuchle D, Larsson J, Kennison JA, Rasmuson-Lestander A, Muller J (2001) Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development 128(17):3371–3379

Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LL, Ito S, Cooper S, Kondo K, Koseki Y, Ishikura T, Long HK, Sheahan TW, Brockdorff N, Kessler BM, Koseki H, Klose RJ (2014) Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell 157(6):1445–1459. https://doi.org/10.1016/j.cell.2014.05.004

Blackledge NP, Fursova NA, Kelley JR, Huseyin MK, Feldmann A, Klose RJ (2020) PRC1 catalytic activity is central to Polycomb system function. Mol cell 77(4):857–874 e859. https://doi.org/10.1016/j.molcel.2019.12.001

Bodega B, Marasca F, Ranzani V, Cherubini A, Della Valle F, Neguembor MV, Wassef M, Zippo A, Lanzuolo C, Pagani M, Orlando V (2017) A cytosolic Ezh1 isoform modulates a PRC2-Ezh1 epigenetic adaptive response in postmitotic cells. Nat Struct Mol Biol 24(5):444–452. https://doi.org/10.1038/nsmb.3392

Boulay G, Rosnoblet C, Guerardel C, Angrand PO, Leprince D (2011) Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. Biochem J 434(2):333–342. https://doi.org/10.1042/BJ20100944

Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K (2003) EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J 22(20):5323–5335. https://doi.org/10.1093/emboj/cdg542

Bratkowski M, Yang X, Liu X (2017) Polycomb repressive complex 2 in an autoinhibited state. J Biol Chem 292(32):13323–13332. https://doi.org/10.1074/jbc.M117.787572

Bratkowski M, Yang X, Liu X (2018) An evolutionarily conserved structural platform for PRC2 inhibition by a class of Ezh2 inhibitors. Sci Rep 8(1):9092. https://doi.org/10.1038/s41598-018-27175-w

Brien GL, Gambero G, O'Connell DJ, Jerman E, Turner SA, Egan CM, Dunne EJ, Jurgens MC, Wynne K, Piao L, Lohan AJ, Ferguson N, Shi X, Sinha KM, Loftus BJ, Cagney G, Bracken AP (2012) Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol 19(12):1273–1281. https://doi.org/10.1038/nsmb.2449

Brockdorff N (2013) Noncoding RNA and Polycomb recruitment. RNA 19(4):429–442. https://doi.org/10.1261/rna.037598.112

Brooun A, Gajiwala KS, Deng YL, Liu W, Bolanos B, Bingham P, He YA, Diehl W, Grable N, Kung PP, Sutton S, Maegley KA, Yu X, Stewart AE (2016) Polycomb repressive complex 2 structure with inhibitor reveals a mechanism of activation and drug resistance. Nat Commun 7:11384. https://doi.org/10.1038/ncomms11384

Brown ZZ, Muller MM, Jain SU, Allis CD, Lewis PW, Muir TW (2014) Strategy for "detoxification" of a cancer-derived histone mutant based on mapping its interaction with the methyltransferase PRC2. J Am Chem Soc 136(39):13498–13501. https://doi.org/10.1021/ja5060934

Cai L, Rothbart SB, Lu R, Xu B, Chen WY, Tripathy A, Rockowitz S, Zheng D, Patel DJ, Allis CD, Strahl BD, Song J, Wang GG (2013) An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Mol Cell 49(3):571–582. https://doi.org/10.1016/j.molcel.2012.11.026

Cao R, Zhang Y (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell 15(1):57–67. https://doi.org/10.1016/j.molcel.2004.06.020

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298(5595):1039–1043. https://doi.org/10.1126/science.1076997

Cao R, Wang H, He J, Erdjument-Bromage H, Tempst P, Zhang Y (2008) Role of hPHF1 in H3K27 methylation and Hox gene silencing. Mol Cell Biol 28(5):1862–1872. https://doi.org/10.1128/MCB.01589-07

Casanova M, Preissner T, Cerase A, Poot R, Yamada D, Li X, Appanah R, Bezstarosti K, Demmers J, Koseki H, Brockdorff N (2011) Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development 138(8):1471–1482. https://doi.org/10.1242/dev.053652

Chammas P, Mocavini I, Di Croce L (2020) Engaging chromatin: PRC2 structure meets function. Br J Cancer 122(3):315–328. https://doi.org/10.1038/s41416-019-0615-2

Chan KM, Fang D, Gan H, Hashizume R, Yu C, Schroeder M, Gupta N, Mueller S, James CD, Jenkins R, Sarkaria J, Zhang Z (2013) The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes & Dev 27(9):985–990. https://doi.org/10.1101/gad.217778.113

Chen H, Rossier C, Antonarakis SE (1996) Cloning of a human homolog of the Drosophila enhancer of zeste gene (EZH2) that maps to chromosome 21q22.2. Genomics 38(1):30–37. https://doi.org/10.1006/geno.1996.0588

Chen S, Jiao L, Shubbar M, Yang X, Liu X (2018) Unique structural platforms of Suz12 dictate distinct classes of PRC2 for chromatin binding. Mol Cell 69(5):840–852 e845. https://doi.org/10.1016/j.molcel.2018.01.039

Chen S, Jiao L, Liu X, Yang X, Liu X (2020) A dimeric structural scaffold for PRC2-PCL targeting to CpG Island chromatin. Mol Cell 77(6):1265–1278 e1267. https://doi.org/10.1016/j.molcel.2019.12.019

Chittock EC, Latwiel S, Miller TC, Muller CW (2017) Molecular architecture of polycomb repressive complexes. Biochem Soc Trans 45(1):193–205. https://doi.org/10.1042/BST20160173

Choi J, Bachmann AL, Tauscher K, Benda C, Fierz B, Muller J (2017) DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. Nat Struct Mol Biol. https://doi.org/10.1038/nsmb.3488

Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E (2012) Molecular architecture of human polycomb repressive complex 2. eLife 1:e00005. https://doi.org/10.7554/elife.00005

Ciferri C, Lander GC, Nogales E (2015) Protein domain mapping by internal labeling and single particle electron microscopy. J Struct Biol 192(2):159–162. https://doi.org/10.1016/j.jsb.2015.09.016

Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT (2014) Regulatory interactions between RNA and polycomb repressive complex 2. Mol Cell 55(2):171–185. https://doi.org/10.1016/j.molcel.2014.05.009

Cloos PA, Christensen J, Agger K, Helin K (2008) Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev 22(9):1115–1140. https://doi.org/10.1101/gad.1652908

Cohen AS, Yap DB, Lewis ME, Chijiwa C, Ramos-Arroyo MA, Tkachenko N, Milano V, Fradin M, McKinnon ML, Townsend KN, Xu J, Van Allen MI, Ross CJ, Dobyns WB, Weaver DD, Gibson WT (2016) Weaver syndrome-associated EZH2 protein variants show impaired histone methyltransferase function in vitro. Hum Mutat 37(3):301–307. https://doi.org/10.1002/humu.22946

Coleman RT, Struhl G (2017) Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science 356(6333). https://doi.org/10.1126/science.aai8236

Collins RE, Tachibana M, Tamaru H, Smith KM, Jia D, Zhang X, Selker EU, Shinkai Y, Cheng X (2005) In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem 280(7):5563–5570. https://doi.org/10.1074/jbc.M410483200

Connolly LR, Smith KM, Freitag M (2013) The Fusarium graminearum histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLoS Genet 9(10):e1003916. https://doi.org/10.1371/journal.pgen.1003916

Conway E, Jerman E, Healy E, Ito S, Holoch D, Oliviero G, Deevy O, Glancy E, Fitzpatrick DJ, Mucha M, Watson A, Rice AM, Chammas P, Huang C, Pratt-Kelly I, Koseki Y, Nakayama M, Ishikura T, Streubel G, Wynne K, Hokamp K, McLysaght A, Ciferri C, Di Croce L, Cagney G, Margueron R, Koseki H, Bracken AP (2018) A family of vertebrate-specific polycombs encoded by the LCOR/LCORL genes balance PRC2 subtype activities. Mol Cell 70(3):408–421 e408. https://doi.org/10.1016/j.molcel.2018.03.005

Cooper S, Dienstbier M, Hassan R, Schermelleh L, Sharif J, Blackledge NP, De Marco V, Elderkin S, Koseki H, Klose R, Heger A, Brockdorff N (2014) Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep 7(5):1456–1470. https://doi.org/10.1016/j.celrep.2014.04.012

Cooper S, Grijzenhout A, Underwood E, Ancelin K, Zhang T, Nesterova TB, Anil-Kirmizitas B, Bassett A, Kooistra SM, Agger K, Helin K, Heard E, Brockdorff N (2016) Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat Commun 7:13661. https://doi.org/10.1038/ncomms13661

Coulson M, Robert S, Eyre HJ, Saint R (1998) The identification and localization of a human gene with sequence similarity to Polycomblike of Drosophila melanogaster. Genomics 48(3):381–383. https://doi.org/10.1006/geno.1997.5201

Cyrus S, Burkardt D, Weaver DD, Gibson WT (2019) PRC2-complex related dysfunction in overgrowth syndromes: a review of EZH2, EED, and SUZ12 and their syndromic phenotypes. Am J Med Genet Part C, Semin Med Genet 181(4):519–531. https://doi.org/10.1002/ajmg.c.31754

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111(2):185–196

Davidovich C, Cech TR (2015) The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA 21(12):2007–2022. https://doi.org/10.1261/rna.053918.115

Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol 20(11):1250–1257. https://doi.org/10.1038/nsmb.2679

Davidovich C, Goodrich KJ, Gooding AR, Cech TR (2014) A dimeric state for PRC2. Nucleic Acids Res 42(14):9236–9248. https://doi.org/10.1093/nar/gku540

Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25(10):1010–1022. https://doi.org/10.1101/gad.2037511

Diehl KL, Ge EJ, Weinberg DN, Jani KS, Allis CD, Muir TW (2019) PRC2 engages a bivalent H3K27M-H3K27me3 dinucleosome inhibitor. Proc Natl Acad Sci USA 116(44):22152–22157. https://doi.org/10.1073/pnas.1911775116

Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol 6(8):227. https://doi.org/10.1186/gb-2005-6-8-227

Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, Strahl BD, Fujimori DG, Yates JR 3rd, Madhani HD (2015) Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell 160(1–2):204–218. https://doi.org/10.1016/j.cell.2014.11.039

Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, Waghorn K, Zoi K, Ross FM, Reiter A, Hochhaus A, Drexler HG, Duncombe A, Cervantes F, Oscier D, Boultwood J, Grand FH, Cross NC (2010) Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet 42(8):722–726. https://doi.org/10.1038/ng.621

Gibaja V, Shen F, Harari J, Korn J, Ruddy D, Saenz-Vash V, Zhai H, Rejtar T, Paris CG, Yu Z, Lira M, King D, Qi W, Keen N, Hassan AQ, Chan HM (2016) Development of secondary mutations in wild-type and mutant EZH2 alleles cooperates to confer resistance to EZH2 inhibitors. Oncogene 35(5):558–566. https://doi.org/10.1038/onc.2015.114

Gibson WT, Hood RL, Zhan SH, Bulman DE, Fejes AP, Moore R, Mungall AJ, Eydoux P, Babul-Hirji R, An J, Marra MA, Consortium FC, Chitayat D, Boycott KM, Weaver DD, Jones SJ (2012) Mutations in EZH2 cause Weaver syndrome. Am J Hum Genet 90(1):110–118. https://doi.org/10.1016/j.ajhg.2011.11.018

Grijzenhout A, Godwin J, Koseki H, Gdula MR, Szumska D, McGouran JF, Bhattacharya S, Kessler BM, Brockdorff N, Cooper S (2016) Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Development 143(15):2716–2723. https://doi.org/10.1242/dev.123935

Grzenda A, Lomberk G, Svingen P, Mathison A, Calvo E, Iovanna J, Xiong Y, Faubion W, Urrutia R (2013) Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. Epigenetics Chromatin 6(1):3. https://doi.org/10.1186/1756-8935-6-3

Hamann BL, Blind RD (2018) Nuclear phosphoinositide regulation of chromatin. J Cell Physiol 233(1):107–123. https://doi.org/10.1002/jcp.25886

Han Z, Xing X, Hu M, Zhang Y, Liu P, Chai J (2007) Structural basis of EZH2 recognition by EED. Structure 15(10):1306–1315. https://doi.org/10.1016/j.str.2007.08.007

Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappsilber J, Lerdrup M, Helin K (2008) A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol 10(11):1291–1300. https://doi.org/10.1038/ncb1787

Hauri S, Comoglio F, Seimiya M, Gerstung M, Glatter T, Hansen K, Aebersold R, Paro R, Gstaiger M, Beisel C (2016) A high-density map for navigating the human polycomb complexome. Cell Rep 17(2):583–595. https://doi.org/10.1016/j.celrep.2016.08.096

He GP, Kim S, Ro HS (1999) Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression. J Biol Chem 274(21):14678–14684. https://doi.org/10.1074/jbc.274.21.14678

He Y, Selvaraju S, Curtin ML, Jakob CG, Zhu H, Comess KM, Shaw B, The J, Lima-Fernandes E, Szewczyk MM, Cheng D, Klinge KL, Li HQ, Pliushchev M, Algire MA, Maag D, Guo J, Dietrich J, Panchal SC, Petros AM, Sweis RF, Torrent M, Bigelow LJ, Senisterra G, Li F, Kennedy S, Wu Q, Osterling DJ, Lindley DJ, Gao W, Galasinski S, Barsyte-Lovejoy D, Vedadi M, Buchanan FG, Arrowsmith CH, Chiang GG, Sun C, Pappano WN (2017) The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. Nat Chem Biol 13(4):389–395. https://doi.org/10.1038/nchembio.2306

Healy E, Mucha M, Glancy E, Fitzpatrick DJ, Conway E, Neikes HK, Monger C, Van Mierlo G, Baltissen MP, Koseki Y, Vermeulen M, Koseki H, Bracken AP (2019) PRC2.1 and PRC2.2 synergize to coordinate H3K27 trimethylation. Mol Cell 76(3):437–452 e436. https://doi.org/10.1016/j.molcel.2019.08.012

Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, Hyman AA, Mann M (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell 163(3):712–723. https://doi.org/10.1016/j.cell.2015.09.053

Hennig L, Derkacheva M (2009) Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet 25(9):414–423. https://doi.org/10.1016/j.tig.2009.07.002

Hobert O, Sures I, Ciossek T, Fuchs M, Ullrich A (1996) Isolation and developmental expression analysis of Enx-1, a novel mouse Polycomb group gene. Mech Dev 55(2):171–184

Hojfeldt JW, Hedehus L, Laugesen A, Tatar T, Wiehle L, Helin K (2019) Non-core subunits of the PRC2 complex are collectively required for its target-site specificity. Mol Cell 76(3):423–436 e423. https://doi.org/10.1016/j.molcel.2019.07.031

Hojfeldt JW, Laugesen A, Willumsen BM, Damhofer H, Hedehus L, Tvardovskiy A, Mohammad F, Jensen ON, Helin K (2018) Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. Nat Struct Mol Biol 25(3):225–232. https://doi.org/10.1038/s41594-018-0036-6

Hubner JM, Muller T, Papageorgiou DN, Mauermann M, Krijgsveld J, Russell RB, Ellison DW, Pfister SM, Pajtler KW, Kool M (2019) EZHIP/CXorf67 mimics K27M mutated oncohistones and functions as an intrinsic inhibitor of PRC2 function in aggressive posterior fossa ependymoma. Neuro-oncology 21(7):878–889. https://doi.org/10.1093/neuonc/noz058

Hunkapiller J, Shen Y, Diaz A, Cagney G, McCleary D, Ramalho-Santos M, Krogan N, Ren B, Song JS, Reiter JF (2012) Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. PLoS Genet 8(3):e1002576. https://doi.org/10.1371/journal.pgen.1002576

Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wuhr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP (2015) The BioPlex network: a systematic exploration of the human interactome. Cell 162(2):425–440. https://doi.org/10.1016/j.cell.2015.06.043

Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, Harper JW (2017) Architecture of the human interactome defines protein communities and disease networks. Nature 545(7655):505–509. https://doi.org/10.1038/nature22366

Iglesias N, Currie MA, Jih G, Paulo JA, Siuti N, Kalocsay M, Gygi SP, Moazed D (2018) Automethylation-induced conformational switch in Clr4 (Suv39h) maintains epigenetic stability. Nature 560(7719):504–508. https://doi.org/10.1038/s41586-018-0398-2

Inouye C, Remondelli P, Karin M, Elledge S (1994) Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: the one hybrid system. DNA Cell Biol 13(7):731–742. https://doi.org/10.1089/dna.1994.13.731

Jain SU, Do TJ, Lund PJ, Rashoff AQ, Diehl KL, Cieslik M, Bajic A, Juretic N, Deshmukh S, Venneti S, Muir TW, Garcia BA, Jabado N, Lewis PW (2019) PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. Nat Commun 10(1):2146. https://doi.org/10.1038/s41467-019-09981-6

Jamieson K, Rountree MR, Lewis ZA, Stajich JE, Selker EU (2013) Regional control of histone H3 lysine 27 methylation in neurospora. Proc Natl Acad Sci USA 110(15):6027–6032. https://doi.org/10.1073/pnas.1303750110

Jermann P, Hoerner L, Burger L, Schubeler D (2014) Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. Proc Natl Acad Sci USA 111(33):E3415–E3421. https://doi.org/10.1073/pnas.1400672111

Jiao L, Liu X (2015) Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science 350(6258):aac4383. https://doi.org/10.1126/science.aac4383

Justin N, Zhang Y, Tarricone C, Martin SR, Chen S, Underwood E, De Marco V, Haire LF, Walker PA, Reinberg D, Wilson JR, Gamblin SJ (2016) Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nat Commun 7:11316. https://doi.org/10.1038/ncomms11316

Kahn TG, Dorafshan E, Schultheis D, Zare A, Stenberg P, Reim I, Pirrotta V, Schwartz YB (2016) Interdependence of PRC1 and PRC2 for recruitment to Polycomb response elements. Nucleic Acids Res 44(21):10132–10149. https://doi.org/10.1093/nar/gkw701

Kalb R, Latwiel S, Baymaz HI, Jansen PW, Muller CW, Vermeulen M, Muller J (2014) Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat Struct Mol Biol 21(6):569–571. https://doi.org/10.1038/nsmb.2833

Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R (2013) PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat Struct Mol Biol 20(11):1258–1264. https://doi.org/10.1038/nsmb.2700

Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D (2014a) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol Cell 53(2):290–300. https://doi.org/10.1016/j.molcel.2013.11.012

Kaneko S, Son J, Bonasio R, Shen SS, Reinberg D (2014b) Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. https://doi.org/10.1101/gad.247940.114

Kasinath V, Faini M, Poepsel S, Reif D, Feng XA, Stjepanovic G, Aebersold R, Nogales E (2018) Structures of human PRC2 with its cofactors AEBP2 and JARID2. Science 359(6378):940–944. https://doi.org/10.1126/science.aar5700

Ketel CS, Andersen EF, Vargas ML, Suh J, Strome S, Simon JA (2005) Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol Cell Biol 25(16):6857–6868. https://doi.org/10.1128/MCB.25.16.6857-6868.2005

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA 106(28):11667–11672. https://doi.org/10.1073/pnas.0904715106

Kim TG, Kraus JC, Chen J, Lee Y (2003) JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. J Biol Chem 278(43):42247–42255. https://doi.org/10.1074/jbc.M307386200

Kim H, Kang K, Kim J (2009) AEBP2 as a potential targeting protein for Polycomb repression complex PRC2. Nucleic Acids Res 37(9):2940–2950. https://doi.org/10.1093/nar/gkp149

Kim W, Bird GH, Neff T, Guo G, Kerenyi MA, Walensky LD, Orkin SH (2013) Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. Nat Chem Biol 9(10):643–650. https://doi.org/10.1038/nchembio.1331

Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, Santanach A, Jansen PW, Di Croce L, Vermeulen M (2016) The dynamic interactome and genomic targets of Polycomb

complexes during stem-cell differentiation. Nat Struct Mol Biol 23(7):682–690. https://doi.org/10.1038/nsmb.3248

Knutson SK, Warholic NM, Wigle TJ, Klaus CR, Allain CJ, Raimondi A, Porter Scott M, Chesworth R, Moyer MP, Copeland RA, Richon VM, Pollock RM, Kuntz KW, Keilhack H (2013) Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. Proc Natl Acad Sci USA 110(19):7922–7927. https://doi.org/10.1073/pnas.1303800110

Koontz JI, Soreng AL, Nucci M, Kuo FC, Pauwels P, van Den Berghe H, Dal Cin P, Fletcher JA, Sklar J (2001) Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. Proc Natl Acad Sci USA 98(11):6348–6353. https://doi.org/10.1073/pnas.101132598

Kutateladze TG (2012) Histone deacetylation: IP4 is an epigenetic coregulator. Nat Chem Biol 8(3):230–231. https://doi.org/10.1038/nchembio.795

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of Zeste protein. Genes Dev 16(22):2893–2905. https://doi.org/10.1101/gad.1035902

Kuzmichev A, Jenuwein T, Tempst P, Reinberg D (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell 14(2):183–193

Kuzmichev A, Margueron R, Vaquero A, Preissner TS, Scher M, Kirmizis A, Ouyang X, Brockdorff N, Abate-Shen C, Farnham P, Reinberg D (2005) Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc Natl Acad Sci USA 102(6):1859–1864. https://doi.org/10.1073/pnas.0409875102

Laible G, Wolf A, Dorn R, Reuter G, Nislow C, Lebersorger A, Popkin D, Pillus L, Jenuwein T (1997) Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at *S. cerevisiae telomeres*. EMBO J 16(11):3219–3232. https://doi.org/10.1093/emboj/16.11.3219

Landeira D, Fisher AG (2011) Inactive yet indispensable: the tale of Jarid2. Trends Cell Biol 21(2):74–80. https://doi.org/10.1016/j.tcb.2010.10.004

Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jorgensen HF, Pereira CF, Leleu M, Piccolo FM, Spivakov M, Brookes E, Pombo A, Fisher C, Skarnes WC, Snoek T, Bezstarosti K, Demmers J, Klose RJ, Casanova M, Tavares L, Brockdorff N, Merkenschlager M, Fisher AG (2010) Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol 12(6):618–624. https://doi.org/10.1038/ncb2065

Laprell F, Finkl K, Muller J (2017) Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. Science 356(6333):85–88. https://doi.org/10.1126/science.aai8266

Laugesen A, Hojfeldt JW, Helin K (2016) Role of the Polycomb repressive complex 2 (PRC2) in transcriptional regulation and cancer. Cold Spring Harb Perspect Med 6(9). https://doi.org/10.1101/cshperspect.a026575

Laugesen A, Hojfeldt JW, Helin K (2019) Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. Mol Cell 74(1):8–18. https://doi.org/10.1016/j.molcel.2019.03.011

Lee CH, Holder M, Grau D, Saldana-Meyer R, Yu JR, Ganai RA, Zhang J, Wang M, LeRoy G, Dobenecker MW, Reinberg D, Armache KJ (2018) Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2. Mol Cell 70(3):435–448 e435. https://doi.org/10.1016/j.molcel.2018.03.019

Lee CH, Yu JR, Granat J, Saldana-Meyer R, Andrade J, LeRoy G, Jin Y, Lund P, Stafford JM, Garcia BA, Ueberheide B, Reinberg D (2019) Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. Genes Dev 33(19–20):1428–1440. https://doi.org/10.1101/gad.328773.119

Lewis EB (1978) A gene complex controlling segmentation in Drosophila. Nature 276(5688):565–570

Lewis PW, Muller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ, Allis CD (2013) Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 340(6134):857–861. https://doi.org/10.1126/science.1232245

Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D (2010) Jarid2 and PRC2, partners in regulating gene expression. Genes Dev 24(4):368–380. https://doi.org/10.1101/gad.1886410

Li X, Isono K, Yamada D, Endo TA, Endoh M, Shinga J, Mizutani-Koseki Y, Otte AP, Casanova M, Kitamura H, Kamijo T, Sharif J, Ohara O, Toyada T, Bernstein BE, Brockdorff N, Koseki H (2011) Mammalian polycomb-like Pcl2/Mtf2 is a novel regulatory component of PRC2 that can differentially modulate polycomb activity both at the Hox gene cluster and at Cdkn2a genes. Mol Cell Biol 31(2):351–364. https://doi.org/10.1128/MCB.00259-10

Li H, Liefke R, Jiang J, Kurland JV, Tian W, Deng P, Zhang W, He Q, Patel DJ, Bulyk ML, Shi Y, Wang Z (2017) Polycomb-like proteins link the PRC2 complex to CpG islands. Nature 549(7671):287–291. https://doi.org/10.1038/nature23881

Liefke R, Shi Y (2015) The PRC2-associated factor C17orf96 is a novel CpG island regulator in mouse ES cells. Cell Discov 1:15008. https://doi.org/10.1038/celldisc.2015.8

Liefke R, Karwacki-Neisius V, Shi Y (2016) EPOP Interacts with Elongin BC and USP7 to modulate the chromatin landscape. Mol Cell 64(4):659–672. https://doi.org/10.1016/j.molcel.2016.10.019

Long Y, Bolanos B, Gong L, Liu W, Goodrich KJ, Yang X, Chen S, Gooding AR, Maegley KA, Gajiwala KS, Brooun A, Cech TR, Liu X (2017) Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. eLife 6. https://doi.org/10.7554/elife.31558

Ma X, Wang J, Wang J, Ma CX, Gao X, Patriub V, Sklar JL (2017) The JAZF1-SUZ12 fusion protein disrupts PRC2 complexes and impairs chromatin repression during human endometrial stromal tumorogenesis. Oncotarget 8(3):4062–4078. https://doi.org/10.18632/oncotarget.13270

Maier VK, Feeney CM, Taylor JE, Creech AL, Qiao JW, Szanto A, Das PP, Chevrier N, Cifuentes-Rojas C, Orkin SH, Carr SA, Jaffe JD, Mertins P, Lee JT (2015) Functional proteomic analysis of repressive histone methyltransferase complexes reveals ZNF518B as a G9A regulator. Mol Cell Proteomics: MCP 14(6):1435–1446. https://doi.org/10.1074/mcp.M114.044586

Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469(7330):343–349. https://doi.org/10.1038/nature09784

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D (2008) Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell 32(4):503–518. https://doi.org/10.1016/j.molcel.2008.11.004

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ 3rd, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461(7265):762–767. https://doi.org/10.1038/nature08398

Martin C, Cao R, Zhang Y (2006) Substrate preferences of the EZH2 histone methyltransferase complex. J Biol Chem 281(13):8365–8370. https://doi.org/10.1074/jbc.M513425200

McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y, Smitheman KN, Ott HM, Pappalardi MB, Allen KE, Chen SB, Della Pietra A, 3rd, Dul E, Hughes AM, Gilbert SA, Thrall SH, Tummino PJ, Kruger RG, Brandt M, Schwartz B, Creasy CL (2012a) Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). Proc Nat Acad Sci USA 109(8):2989–2994. https://doi.org/10.1073/pnas.1116418109

McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A, 3rd, Diaz E, LaFrance LV, Mellinger M, Duquenne C, Tian X, Kruger RG, McHugh CF, Brandt M, Miller WH, Dhanak D, Verma SK, Tummino PJ, Creasy CL (2012b) EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature 492(7427):108–112. https://doi.org/10.1038/nature11606

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet 6(12):e1001244. https://doi.org/10.1371/journal.pgen.1001244

Min J, Zhang X, Cheng X, Grewal SI, Xu RM (2002) Structure of the SET domain histone lysine methyltransferase Clr4. Nat Struct Biol 9(11):828–832. https://doi.org/10.1038/nsb860

Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, Shah S, Zhu H, Kimbara M, Shashkin P, Charlot JF, Tcherpakov M, Corbett R, Tam A, Varhol R, Smailus D, Moksa M, Zhao Y, Delaney A, Qian H, Birol I, Schein J, Moore R, Holt R, Horsman DE, Connors JM, Jones S, Aparicio S, Hirst M, Gascoyne RD, Marra MA (2010) Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 42(2):181–185. https://doi.org/10.1038/ng.518

Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, Margueron R, Ait-Si-Ali S (2014) The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol Cell 53(2):277–289. https://doi.org/10.1016/j.molcel.2013.12.005

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 111(2):197–208

Murzina NV, Pei XY, Zhang W, Sparkes M, Vicente-Garcia J, Pratap JV, McLaughlin SH, Ben-Shahar TR, Verreault A, Luisi BF, Laue ED (2008) Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure 16(7):1077–1085. https://doi.org/10.1016/j.str.2008.05.006

Musselman CA, Avvakumov N, Watanabe R, Abraham CG, Lalonde ME, Hong Z, Allen C, Roy S, Nunez JK, Nickoloff J, Kulesza CA, Yasui A, Cote J, Kutateladze TG (2012) Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol 19(12):1266–1272. https://doi.org/10.1038/nsmb.2435

Nalefski EA, Falke JJ (1996) The C2 domain calcium-binding motif: structural and functional diversity. Protein Sci Publ Protein Soc 5(12):2375–2390. https://doi.org/10.1002/pro.5560051201

Nekrasov M, Wild B, Muller J (2005) Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep 6(4):348–353. https://doi.org/10.1038/sj.embor.7400376

O'Connell S, Wang L, Robert S, Jones CA, Saint R, Jones RS (2001) Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. J Biol Chem 276(46):43065–43073. https://doi.org/10.1074/jbc.M104294200

Oksuz O, Narendra V, Lee CH, Descostes N, LeRoy G, Raviram R, Blumenberg L, Karch K, Rocha PP, Garcia BA, Skok JA, Reinberg D (2018) Capturing the Onset of PRC2-mediated repressive domain formation. Mol Cell 70(6):1149–1162 e1145. https://doi.org/10.1016/j.molcel.2018.05.023

Oliviero G, Brien GL, Waston A, Streubel G, Jerman E, Andrews D, Doyle B, Munawar N, Wynne K, Crean J, Bracken AP, Cagney G (2016) Dynamic protein interactions of the polycomb repressive complex 2 during differentiation of pluripotent cells. Mol Cell Proteomics MCP 15(11):3450–3460. https://doi.org/10.1074/mcp.M116.062240

Pajtler KW, Wen J, Sill M, Lin T, Orisme W, Tang B, Hubner JM, Ramaswamy V, Jia S, Dalton JD, Haupfear K, Rogers HA, Punchihewa C, Lee R, Easton J, Wu G, Ritzmann TA, Chapman R, Chavez L, Boop FA, Klimo P, Sabin ND, Ogg R, Mack SC, Freibaum BD, Kim HJ, Witt H, Jones DTW, Vo B, Gajjar A, Pounds S, Onar-Thomas A, Roussel MF, Zhang J, Taylor JP, Merchant TE, Grundy R, Tatevossian RG, Taylor MD, Pfister SM, Korshunov A, Kool M, Ellison DW (2018) Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. Acta Neuropathol 136(2):211–226. https://doi.org/10.1007/s00401-018-1877-0

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32(2):232–246. https://doi.org/10.1016/j.molcel.2008.08.022

Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 464(7286):306–310. https://doi.org/10.1038/nature08788

Patsialou A, Wilsker D, Moran E (2005) DNA-binding properties of ARID family proteins. Nucleic Acids Res 33(1):66–80. https://doi.org/10.1093/nar/gki145

Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 139(7):1290–1302. https://doi.org/10.1016/j.cell.2009.12.002

Perino M, van Mierlo G, Karemaker ID, van Genesen S, Vermeulen M, Marks H, van Heeringen SJ, Veenstra GJC (2018) MTF2 recruits polycomb repressive complex 2 by helical-shape-selective DNA binding. Nat Genet 50(7):1002–1010. https://doi.org/10.1038/s41588-018-0134-8

Piunti A, Smith ER, Morgan MAJ, Ugarenko M, Khaltyan N, Helmin KA, Ryan CA, Murray DC, Rickels RA, Yilmaz BD, Rendleman EJ, Savas JN, Singer BD, Bulun SE, Shilatifard A (2019) CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. Sci Adv 5(7):eaax2887. https://doi.org/10.1126/sciadv.aax2887

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300(5616):131–135. https://doi.org/10.1126/science.1084274

Poepsel S, Kasinath V, Nogales E (2018) Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nat Struct Mol Biol 25(2):154–162. https://doi.org/10.1038/s41594-018-0023-y

Qi W, Zhao K, Gu J, Huang Y, Wang Y, Zhang H, Zhang M, Zhang J, Yu Z, Li L, Teng L, Chuai S, Zhang C, Zhao M, Chan H, Chen Z, Fang D, Fei Q, Feng L, Feng L, Gao Y, Ge H, Ge X, Li G, Lingel A, Lin Y, Liu Y, Luo F, Shi M, Wang L, Wang Z, Yu Y, Zeng J, Zeng C, Zhang L, Zhang Q, Zhou S, Oyang C, Atadja P, Li E (2017) An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. Nat Chem Biol 13(4):381–388. https://doi.org/10.1038/nchembio.2304

Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, Behrens A, Bertone P, Hendrich B (2012) NuRD-mediated deacetylation of H3K27 facilitates recruitment of polycomb repressive complex 2 to direct gene repression. EMBO J 31(3):593–605. https://doi.org/10.1038/emboj.2011.431

Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K (2014) Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell 55(3):347–360. https://doi.org/10.1016/j.molcel.2014.06.005

Ringrose L (2017) Noncoding RNAs in polycomb and trithorax regulation: a quantitative perspective. Annu Rev Genet 51:385–411. https://doi.org/10.1146/annurev-genet-120116-023402

Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134(2):223–232. https://doi.org/10.1242/dev.02723

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311–1323. https://doi.org/10.1016/j.cell.2007.05.022

Sanulli S, Justin N, Teissandier A, Ancelin K, Portoso M, Caron M, Michaud A, Lombard B, da Rocha ST, Offer J, Loew D, Servant N, Wassef M, Burlina F, Gamblin SJ, Heard E, Margueron R (2015) Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation. Mol Cell 57(5):769–783. https://doi.org/10.1016/j.molcel.2014.12.020

Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D (2008) Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol Cell Biol 28(8):2718–2731. https://doi.org/10.1128/MCB.02017-07

Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Muller J (2010) Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature 465(7295):243–247. https://doi.org/10.1038/nature08966

Schmitges FW, Prusty AB, Faty M, Stutzer A, Lingaraju GM, Aiwazian J, Sack R, Hess D, Li L, Zhou S, Bunker RD, Wirth U, Bouwmeester T, Bauer A, Ly-Hartig N, Zhao K, Chan H, Gu J, Gut H, Fischle W, Muller J, Thoma NH (2011) Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell 42(3):330–341. https://doi.org/10.1016/j.molcel.2011.03.025

Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G (2017) Genome regulation by polycomb and trithorax: 70 years and counting. Cell 171(1):34–57. https://doi.org/10.1016/j.cell.2017.08.002

Schwartz YB, Pirrotta V (2013) A new world of Polycombs: unexpected partnerships and emerging functions. Nat Rev Genet 14(12):853–864. https://doi.org/10.1038/nrg3603

Sewalt RG, van der Vlag J, Gunster MJ, Hamer KM, den Blaauwen JL, Satijn DP, Hendrix T, van Driel R, Otte AP (1998) Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. Mol Cell Biol 18(6):3586–3595

Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X, Yuan GC, Orkin SH (2008) EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell 32(4):491–502. https://doi.org/10.1016/j.molcel.2008.10.016

Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH (2009) Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 139(7):1303–1314. https://doi.org/10.1016/j.cell.2009.12.003

Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AH, Jenuwein T, Otte AP, Brockdorff N (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell 4(4):481–495

Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10(10):697–708. https://doi.org/10.1038/nrm2763

Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res 41(1):e28. https://doi.org/10.1093/nar/gks941

Sneeringer CJ, Scott MP, Kuntz KW, Knutson SK, Pollock RM, Richon VM, Copeland RA (2010) Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci USA 107(49):20980–20985. https://doi.org/10.1073/pnas.1012525107

Son J, Shen SS, Margueron R, Reinberg D (2013) Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev 27(24):2663–2677. https://doi.org/10.1101/gad.225888.113

Song JJ, Garlick JD, Kingston RE (2008) Structural basis of histone H4 recognition by p55. Genes Dev 22(10):1313–1318. https://doi.org/10.1101/gad.1653308

Stafford JM, Lee CH, Voigt P, Descostes N, Saldana-Meyer R, Yu JR, Leroy G, Oksuz O, Chapman JR, Suarez F, Modrek AS, Bayin NS, Placantonakis DG, Karajannis MA, Snuderl M, Ueberheide B, Reinberg D (2018) Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma. Sci Adv 4(10):eaau5935. https://doi.org/10.1126/sciadv.aau5935

Stirnimann CU, Petsalaki E, Russell RB, Muller CW (2010) WD40 proteins propel cellular networks. Trends Biochem Sci 35(10):565–574. https://doi.org/10.1016/j.tibs.2010.04.003

Sun A, Li F, Liu Z, Jiang Y, Zhang J, Wu J, Shi Y (2018) Structural and biochemical insights into human zinc finger protein AEBP2 reveals interactions with RBBP4. Protein Cell 9(8):738–742. https://doi.org/10.1007/s13238-017-0483-6

Takeuchi T, Yamazaki Y, Katoh-Fukui Y, Tsuchiya R, Kondo S, Motoyama J, Higashinakagawa T (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes Dev 9(10):1211–1222. https://doi.org/10.1101/gad.9.10.1211

Tamburri S, Lavarone E, Fernandez-Perez D, Conway E, Zanotti M, Manganaro D, Pasini D (2020) Histone H2AK119 Mono-ubiquitination is essential for polycomb-mediated transcriptional repression. Mol Cell 77(4):840–856 e845. https://doi.org/10.1016/j.molcel.2019.11.021

Tatton-Brown K, Hanks S, Ruark E, Zachariou A, Duarte Sdel V, Ramsay E, Snape K, Murray A, Perdeaux ER, Seal S, Loveday C, Banka S, Clericuzio C, Flinter F, Magee A, McConnell V, Patton M, Raith W, Rankin J, Splitt M, Strenger V, Taylor C, Wheeler P, Temple KI, Cole T, Childhood Overgrowth C, Douglas J, Rahman N (2011) Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. Oncotarget 2(12):1127–1133. https://doi.org/10.18632/oncotarget.385

Thompson PR, Wang D, Wang L, Fulco M, Pediconi N, Zhang D, An W, Ge Q, Roeder RG, Wong J, Levrero M, Sartorelli V, Cotter RJ, Cole PA (2004) Regulation of the p300 HAT domain via a novel activation loop. Nat Struct Mol Biol 11(4):308–315. https://doi.org/10.1038/nsmb740

Tie F, Prasad-Sinha J, Birve A, Rasmuson-Lestander A, Harte PJ (2003) A 1-megadalton ESC/E(Z) complex from Drosophila that contains polycomblike and RPD3. Mol Cell Biol 23(9):3352–3362

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329(5992):689–693. https://doi.org/10.1126/science.1192002

Tsukiyama T, Wu C (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell 83(6):1011–1020. https://doi.org/10.1016/0092-8674(95)90216-3

van Lohuizen M, Tijms M, Voncken JW, Schumacher A, Magnuson T, Wientjens E (1998) Interaction of mouse polycomb-group (Pc-G) proteins Enx1 and Enx2 with Eed: indication for separate Pc-G complexes. Mol Cell Biol 18(6):3572–3579

van Mierlo G, Veenstra GJC, Vermeulen M, Marks H (2019) The complexity of PRC2 subcomplexes. Trends Cell Biol 29(8):660–671. https://doi.org/10.1016/j.tcb.2019.05.004

Vaswani RG, Gehling VS, Dakin LA, Cook AS, Nasveschuk CG, Duplessis M, Iyer P, Balasubramanian S, Zhao F, Good AC, Campbell R, Lee C, Cantone N, Cummings RT, Normant E, Bellon SF, Albrecht BK, Harmange JC, Trojer P, Audia JE, Zhang Y, Justin N, Chen S, Wilson JR, Gamblin SJ (2016) Identification of (R)-N-((4-Methoxy-6-methyl-2-oxo-1,2-dihydropyridin-3-yl)methyl)-2-methyl-1-(1-(1 -(2,2,2-trifluoroethyl)piperidin-4-yl)ethyl)-1H-indole-3-carboxamide (CPI-1205), a potent and selective inhibitor of histone methyltransferase EZH2, suitable for phase I clinical trials for B-cell lymphomas. J Med Chem 59(21):9928–9941. https://doi.org/10.1021/acs.jmedchem.6b01315

Verdine GL, Hilinski GJ (2012) Stapled peptides for intracellular drug targets. Methods Enzymol 503:3–33. https://doi.org/10.1016/B978-0-12-396962-0.00001-X

Vizan P, Beringer M, Ballare C, Di Croce L (2015) Role of PRC2-associated factors in stem cells and disease. FEBS J 282(9):1723–1735. https://doi.org/10.1111/febs.13083

Wachter E, Quante T, Merusi C, Arczewska A, Stewart F, Webb S, Bird A (2014) Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. eLife 3:e03397. https://doi.org/10.7554/elife.03397

Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K, Torchia J, Krogan NJ, Reiter JF, Stanford WL (2010) Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. Cell Stem Cell 6(2):153–166. https://doi.org/10.1016/j.stem.2009.12.014

Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y (2004a) Role of histone H2A ubiquitination in Polycomb silencing. Nature 431(7010):873–878. https://doi.org/10.1038/nature02985

Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, Jones RS (2004b) Hierarchical recruitment of polycomb group silencing complexes. Mol Cell 14(5):637–646. https://doi.org/10.1016/j.molcel.2004.05.009

Wang S, Robertson GP, Zhu J (2004c) A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers. Gene 343(1):69–78. https://doi.org/10.1016/j.gene.2004.09.006

Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, Youmans DT, Cech TR, Davidovich C (2017a) Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. Mol Cell 65(6):1056–1067 e1055. https://doi.org/10.1016/j.molcel.2017.02.003

Wang X, Paucek RD, Gooding AR, Brown ZZ, Ge EJ, Muir TW, Cech TR (2017b) Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nat Struct Mol Biol. https://doi.org/10.1038/nsmb.3487

Wang X, Long Y, Paucek RD, Gooding AR, Lee T, Burdorf RM, Cech TR (2019) Regulation of histone methylation by automethylation of PRC2. Genes Dev 33(19–20):1416–1427. https://doi.org/10.1101/gad.328849.119

Watson PJ, Fairall L, Santos GM, Schwabe JW (2012) Structure of HDAC3 bound to co-repressor and inositol tetraphosphate. Nature 481(7381):335–340. https://doi.org/10.1038/nature10728

Whittaker C, Dean C (2017) The FLC locus: a platform for discoveries in epigenetics and adaptation. Annu Rev Cell Dev Biol 33:555–575. https://doi.org/10.1146/annurev-cellbio-100616-060546

Wu H, Zeng H, Dong A, Li F, He H, Senisterra G, Seitova A, Duan S, Brown PJ, Vedadi M, Arrowsmith CH, Schapira M (2013) Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. PLoS ONE 8(12):e83737. https://doi.org/10.1371/journal.pone.0083737

Wutz A (2011) Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet 12(8):542–553. https://doi.org/10.1038/nrg3035

Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell S, Taylor IA, Blackburn GM, Gamblin SJ (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature 421(6923):652–656. https://doi.org/10.1038/nature01378

Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, Qiu W, Liu H, Jones AE, MacKenzie F, Pan P, Li SS, Wang H, Min J (2010) Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc Natl Acad Sci USA 107(45):19266–19271. https://doi.org/10.1073/pnas.1008937107

Yamamoto K, Sonoda M, Inokuchi J, Shirasawa S, Sasazuki T (2004) Polycomb group suppressor of zeste 12 links heterochromatin protein 1alpha and enhancer of zeste 2. J Biol Chem 279(1):401–406. https://doi.org/10.1074/jbc.M307344200

Youmans DT, Schmidt JC, Cech TR (2018) Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. Genes Dev 32(11–12):794–805. https://doi.org/10.1101/gad.311936.118

Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D (1997) Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell 89(3):357–364

Zhang Y, LeRoy G, Seelig HP, Lane WS, Reinberg D (1998) The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell 95(2):279–289. https://doi.org/10.1016/s0092-8674(00)81758-4

Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Selker EU, Cheng X (2002) Structure of the neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. Cell 111(1):117–127. https://doi.org/10.1016/s0092-8674(02)00999-6

Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, Selker EU, Cheng X (2003) Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell 12(1):177–185. https://doi.org/10.1016/s1097-2765(03)00224-7

Zhang Z, Jones A, Sun CW, Li C, Chang CW, Joo HY, Dai Q, Mysliwiec MR, Wu LC, Guo Y, Yang W, Liu K, Pawlik KM, Erdjument-Bromage H, Tempst P, Lee Y, Min J, Townes TM, Wang H (2011) PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. Stem Cells 29(2):229–240. https://doi.org/10.1002/stem.578

Zhang Q, McKenzie NJ, Warneford-Thomson R, Gail EH, Flanigan SF, Owen BM, Lauman R, Levina V, Garcia BA, Schittenhelm RB, Bonasio R, Davidovich C (2019) RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. Nat Struct Mol Biol 26(3):237–247. https://doi.org/10.1038/s41594-019-0197-y

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322(5902):750–756. https://doi.org/10.1126/science.1163045

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40(6):939–953. https://doi.org/10.1016/j.molcel.2010.12.011

Chapter 18
Assembly and Function of the Anthrax Toxin Protein Translocation Complex

Robert C. Liddington

Abstract Anthrax toxin is a major virulence factor of *Bacillus anthracis*, a Gram-positive bacterium which can form highly stable spores that are the causative agents of the disease, anthrax. While chiefly a disease of livestock, spores can be "weaponized" as a bio-terrorist agent, and can be deadly if not recognized and treated early with antibiotics. The intracellular pathways affected by the enzymes are broadly understood and are not discussed here. This chapter focuses on what is known about the assembly of secreted toxins on the host cell surface and how the toxin is delivered into the cytosol. The central component is the "Protective Antigen", which self-oligomerizes and forms complexes with its pay-load, either Lethal Factor or Edema Factor. It binds a host receptor, CMG2, or a close relative, triggering receptor-mediated endocytosis, and forms a remarkably elegant yet powerful machine that delivers toxic enzymes into the cytosol, powered only by the pH gradient across the membrane. We now have atomic structures of most of the starting, intermediate and final assemblies in the infectious process. Together with a major body of biophysical, mutational and biochemical work, these studies reveal a remarkable story of both how toxin assembly is choreographed in time and space.

Keywords Anthrax · Pathogenesis · Structure · Macromolecular assembly · Conformational change · Protein translocation · pH trigger

Introduction

Anthrax toxin is a major virulence factor of *Bacillus anthracis*, a Gram-positive, non-motile, rod-shaped, encapsulated bacterium, which can also form highly stable spores that are the causative agents of the disease, anthrax. Disease begins by attachment of spores to mucosa-associated lymphoid tissues of mammals, enabling spore germination and the formation of replication-competent "vegetative cells", initially replicating in the lymphatic system, before migrating to the vasculature, ultimately

R. C. Liddington (✉)
7602 ½ Eads Ave., La Jolla, CA 92037, USA
e-mail: btblidders@gmail.com

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes III: Structure and Function*, Subcellular Biochemistry 96,
https://doi.org/10.1007/978-3-030-58971-4_18

reaching very high levels that are often lethal if not diagnosed at an early stage (reviewed in Moayeri and Leppla 2004, 2009; Frankel et al. 2009). While chiefly a disease of livestock, spores can be "weaponized" as a bio-terrorist agent.

Two extrachromosomal plasmids, pX01 and pX02, encode the major pathogenic determinants of *Bacillus anthracis*—the tripartite Anthrax Toxin (Larson et al. 1988) and a poly(γ-D)glutamate capsule—which distinguish it from its harmless relative, *Bacillus cereus*. The capsule acts as a physical barrier as well as a "cloaking device" against immune surveillance [reviewed in (Moayeri et al. 2015)]. Anthrax Toxin plays critical roles in both early (stifling the innate immune response) and late (destruction of the host vasculature) stages of infection. The intracellular pathways affected by the enzymes are broadly understood, although the details are dependent on host/strain, cell type and disease progression; these are discussed elsewhere [see (Moayeri et al. 2015) and references therein]. This chapter focuses on what we know about the assembly of secreted toxins on the host cell surface and how the toxin is delivered into the cytosol.

The central component of Anthrax Toxin is the "Protective Antigen" (PA) (so-called because it is the major component of anthrax vaccines Gu et al. 1999; Leppla et al. 2002), which self-oligomerizes and forms complexes with its pay-load, either or both of two secreted toxic enzymes, the Lethal Factor (LF) and Edema Factor (EF), all encoded on the pX01 plasmid (Singh et al. 1999). PA also binds its host cell receptor, CMG2 (Bradley et al. 2001) or a close relative (van der Goot and Young 2009), triggering clathrin-dependent receptor-mediated endocytic (Abrami et al. 2003), and ultimately forms a remarkably elegant yet powerful machine that delivers the toxic enzymes into the cytosol, powered only by the pH gradient across the endosomal-cytosolic membrane (reviewed in Young and Collier 2007).

We now have crystal and EM structures of nearly all of the starting, intermediate and final structures in the infectious process, most at atomic or near-atomic resolution. Together with a major body of biophysical, mutational and biochemical work, these studies reveal a remarkable story of both how the toxin assembles, and how assembly, and thus, pathogenesis, may be choreographed in time and space. Here, I will focus on the facts, but identify features of the story that remain unresolved or ambiguous, and which could form the basis for future research.

Anthrax Protective Antigen and Prepore Formation

PA is secreted by *B. anthracis* as a water-soluble 83 kDa monomeric pro-protein, "$PA_{\mathbf{83}}$". It is organized into 4 domains ("D1"–"D4") (Petosa et al. 1997) (Fig. 18.1), and is fully competent to bind its host receptor, CMG2 (*see* below). However, $PA_{\mathbf{83}}$ cannot oligomerize to form a water-soluble *prepore* owing to steric hindrance between N-terminal segments of D1 (Petosa et al. 1997). Instead, a host cell-surface furin-like protease must first cleave a loop between two β-strands within the major sheet of D1 (after residue Arg262), creating N-terminal ($PA_{\mathbf{21}}$) and C-terminal ($PA_{\mathbf{63}}$) fragments (Molloy et al. 1992; Gordon et al. 1995). But physical separation of these

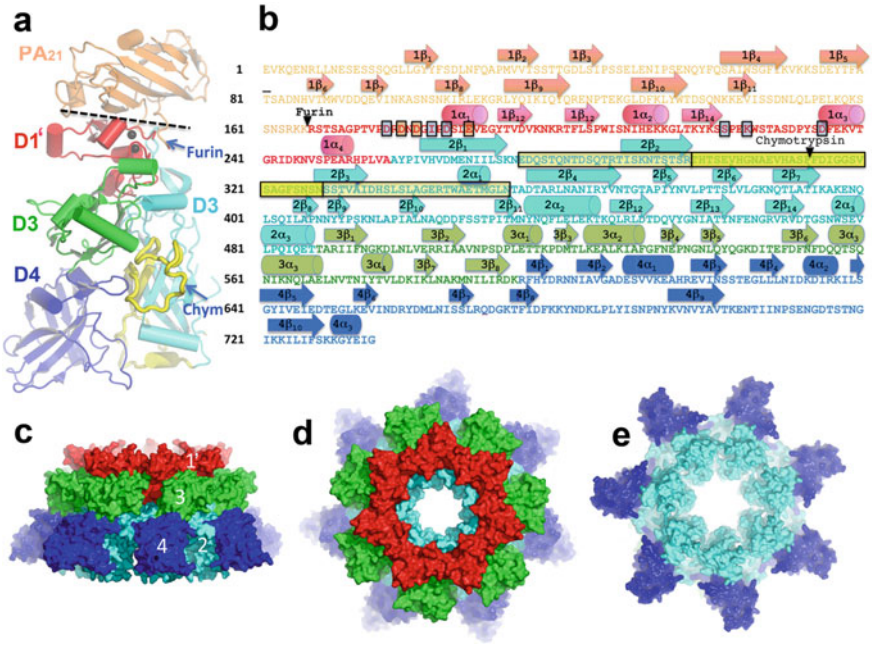


Fig. 18.1 Structure and sequence of PA in its monomeric and prepore forms. **a**. Structure of full-length PA, with secondary structure colored by domain; except for D2, for which the residues destined to form the central barrel of the pore are colored yellow. A black dashed line indicates where the main sheet of D1 must be torn apart to create PA_{21} and PA**63** following cleavage or "nicking" by furin. The two black balls are Ca^{2+} ions. **b**. Sequence of PA**83** colored as in **a**, with secondary structure elements labeled. Residues coordinating the two Ca^{2+} ions are highlighted in rectangular boxes; non-acidic residues coordinate via their main-chain carbonyls. Orange highlights indicate residues that bridge both Ca^{2+} ions. **c**-**e** Surface representations of the heptameric prepore, colored as in **a**. **c** is a side view showing that D1′ is membrane distal and D4 membrane proximal. **d** is a "top" view revealing the annulus formed by interlocking D1′ domains. D2 is just evident along the channel lumen. E. is a view form the bottom, i.e. the membrane-proximal face. The surface comprises exclusively D2 domains lining the channel lumen, and D4 domains on the periphery

fragments requires rupture of the major β-sheet of D1. This process is slow in vitro, but can be achieved by passage of "nicked" $PA_{\mathbf{83}}$ over an ion-exchange column, in which case a heptameric species $(PA\mathbf{63})_7$, is the principal product. However, when an excess of LF or EF is included, a functional octameric species, $(PA\mathbf{63})_8$, predominates (Kintzer et al. 2009), that may play a dominant role at late stages of infection. Crystal structures of both oligomers have been determined, and they are functionally and structurally very similar (the difference in packing angle between PA_{63} protomers in the two oligomers is only ~6°, which can be readily accommodated by small sidechain adjustments), and for the sake of brevity, the following discussion will generally refer to the heptameric species.

Prepore Structure

A low resolution (4.5 Å resolution) crystal structure of the heptameric $(PA_{63})_7$ *prepore* was first described in 1997 (Petosa et al. 1997) (Fig. 18.1c–e). Despite the limited resolution, unbiased phasing (utilizing the heptamer's non-crystallographic symmetry) demonstrated convincingly that oligomerization comprised essentially a rigid body assembly of $PA_{\mathbf{63}}$ protomers, with a modest shift in one surface loop implicated in LF binding. Later high-resolution (2.8 Å) structures of the heptamer (Lacy et al. 2004) and octamer (Feld et al. 2010) confirmed these features.

The $PA_{\mathbf{63}}$ protomers lie with their longest axes running roughly parallel to the heptamer axis (Fig. 18.1c). The "upper" (*i.e.,* membrane-distal) part of the heptamer is formed by tight self-association of 7 D1′ domains, which form an annulus that defines the top of the prepore (Fig. 18.1d). D3 domains lie just beneath the D1′ domains; they make no contact with each other, but insert one helix (residues 512–517) into the D1′ domain of their counterclockwise neighbor, propping up its surface-exposed β-hairpin (1β**12**–1β**13**) by ~5 Å—the largest conformational change upon heptamer formation. D1′ is also stabilized by a pair of Ca^{2+} ions which share 3 coordinating Asp/Glu ligands, lying just beneath the newly exposed cleaved face of D1′, where they presumably help to stabilize the vestigial "half-domains" (Fig. 18.1a). The D2 domains pack around the heptamer axis and extend along most of its length, tightly to each other (and to the D1′ domains) at the top, but more loosely at the base. The prepore lumen has a tightest constriction of ~20 Å in diameter, near the top of the D2 domains. At the base of heptamer, only 2 domains form the lower (membrane-proximal) surface: the D2 annulus near the center and the D4 domains, which lie distal to the axis. D4 domains make no contacts with each other, but make several important contacts with D2 of their own protomer (*see* below), as well as forming a cage with D2 of the clockwise protomer which provides a safe haven for the extended protease-sensitive pro-hairpin loop of D2 (Fig. 18.1a), This protection may be critical, since cleavage of one or more of these 7 loops enables prepore but not functional pore formation (Singh et al. 1994).

Nature and Role of the Host Receptor, CMG2

CMG2 is abundant on the surfaces of cells of the innate immune system, as well as endothelial cells lining blood vessel walls, and it is highly conserved among the mammalian targets of *B. anthracis.* CMG2 has two extracellular domains of similar size, and PA engages the second (membrane distal) of these, one that is highly homologous to integrin "I domains". Crystal structures of monomeric PA**83** bound to a single copy of the CMG2 "I domain" (Santelli et al. 2004) as well as a lower resolution structure of heptameric $(PA\mathbf{63})_7$ bound to 7 molecules of CMG2 display identical interactions (see Fig. 18.6). CMG2 forms a metal-mediated bond to an Asp residue on D4 via its "MIDAS" motif (Emsley et al. 2000), providing a remarkable

mimic of an integrin-matrix complex. The interaction site on PA D4 was predicted by mutagenesis (Varughese et al. 1999; Rosovitz et al. 2003); this domain is the most variable of the 4 domains in PA orthologs from other species (which presumably utilize different receptors) (Figs. 18.2, 18.3, 18.4 and 18.5).

The bigger surprise was that CMG2 also engages domain D2 of PA, in a region destined to become (the upper) part of the membrane-spanning β-barrel. Thus, by stabilizing this region of D2, as well as cross-linking it to D4, CMG2 acts as a clamp that inhibits conformational changes required for prepore to pore conversion. Indeed, the pH threshold for membrane insertion is reduced from pH 6.0 to 5.0 in the presence of CMG2, suggesting that the receptor discourages premature insertion of the *pore* into the plasma (Abrami et al. 2004). Pore formation occurs at an earlier point during endosome maturation than protein translocation, which occurs only in late endocytic compartments (Abrami et al. 2004), consistent with the lower pH requirement for the second process. That CMG2 remains bound to PA through the steps of receptor-mediated endocytosis, pore formation and protein translocation (Pilpa et al. 2011), plays a role in the timing of pore formation, and may provide a physical support that holds pore at the correct height for membrane insertion, suggests that it should be considered an integral part of toxin translocation machine.

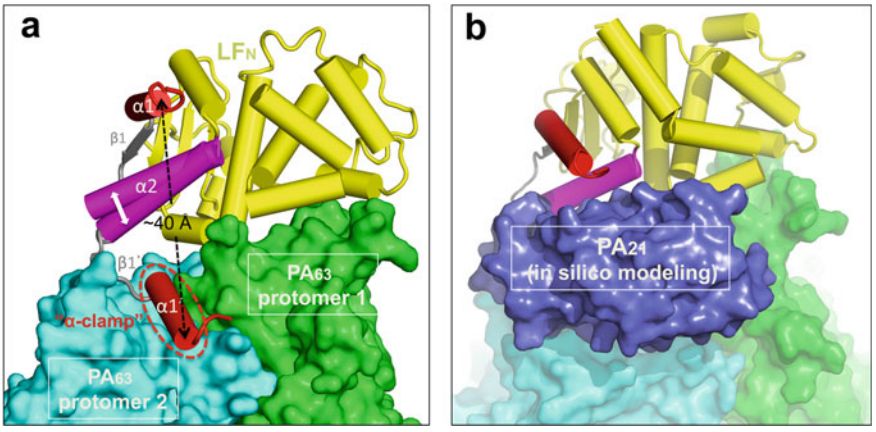


Fig. 18.2 LFN bound to a dimeric PA species. **a**. Two PA protomers are shown as green and cyan surfaces, LFN as yellow cartoon except for the N-terminal segments. In the crystal structure of LF_N bound to oligomeric PA_{63}, the α1 helix undergoes a dramatic shift (~40 Å) compared with the structure of the free enzyme, settling into a shallow pocket called the "α-clamp". The unraveling of the β1 strand and tilting of the α2 helix are consequences of this movement, which positions α1 ~45 Å above the region that will become the Φ-clamp. There are no other significant differences. **b**. A hypothetical model generated in silico in which the PA_{21} element of the cyan protomer (counterclockwise from above) has been added by simple overlay of full-length PA onto PA_{63}. PA**21** binds without major steric clashes, but only to the "native" conformation of LF**N**; that is, PA**21** binding is predicted to inhibit the first step in LF**N** activation

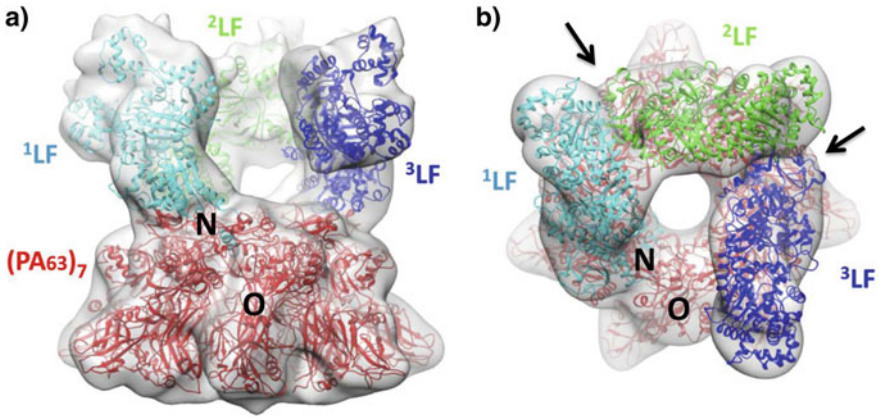


Fig. 18.3 Three molecules of full-length LF bound to the PA**63** heptamer, derived by cryoEM. **a** and **b** Orthogonal views of the hetero-oligomer, with atomic models derived by crystallography fitted into the EM envelope. **a** The "top" view illustrates that the C-terminal domains are suspended up to 30 Å above the heptamer surface, creating large holes in the complex. **b** The "side" view illustrates contact points between the N- and C-termini of consecutive LF molecules which may stabilize the heptamer (shown by black arrows). Since LF binds to two consecutive PA protomers, the odd number of protomers creates a "free" N-terminus (N) and an "orphan" PA protomer (O)

Loading of the Toxic Enzymes and Insights into Prepore Assembly

A high-resolution crystal structure of the octameric prepore bound to four copies of the N-terminal domain of LF (LF_N) has been obtained (Feld et al. 2010), as well as a lower resolution EM structure of the heptameric prepore bound to 3 copies of full-length LF (Fabre et al. 2016). These stoichiometries are consistent with prior mutational and biochemical studies by Collier and co-workers, which showed that the footprint of each LF molecule spans two adjacent protomers of PA (Mogridge et al. 2002). Moreover, their experiments demonstrated that LF did not bind stably to a mutant "nicked" $PA_{\mathbf{83}}$ that could only form monomers; that nicked $PA_{\mathbf{83}}$ did not form dimers; but that mutant pairs with the potential to form only dimers formed a stable complex, a ternary $(PA_{\mathbf{63}})_2LF$ entity. In a separate set of experiments, it was concluded that PA oligomerization is driven by dimeric (toxin-laden) receptor intermediates on the cell surface (Kintzer et al. 2010). So if $(CMG2)_2(PA\mathbf{63})_2LF$ is indeed a stable assembly intermediate for wild-type PA, then it immediately suggests how loading of toxic enzymes onto PA is linked to cell entry. In this scenario, a PA**83** monomer binds a receptor and is then nicked by a cell-surface protease; this can now binds LF, which promotes dimerization with a second PA-laden receptor, which in turn promotes oligomer formation and endocytosis.

In fact, the crystal structures reveal that the footprints of LF and PA_{21} on PA_{63} are almost distinct (i.e., there is only a small degree of overlap between them), which would provide a low energy kinetic pathway for LF to rapidly displace PA**21**,

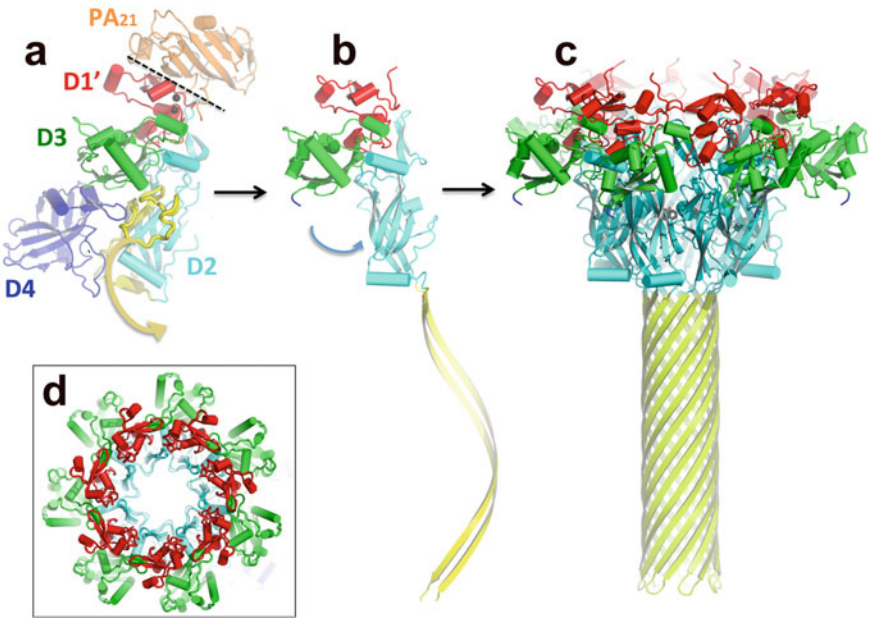


Fig. 18.4 Overview of the transition from monomeric PA to the pore conformation. **a** Monomeric PA oriented so that the invariant D1′ and D3 match those in the pore conformation. An arrows indicates a major point of conformational change: the unraveling of the long b-hairpin (in yellow). **b** The monomeric pore conformation. The arrow indicates the second point of change: the rotation of the D2 barrel, such that it adopts an upright position, enabling seven protomers to pack tightly across the symmetry axis. D4 is not shown, because it is mobile once the binds to D2 have been broken. **c** The result of these two reorganizations in the context of the heptamer is the pore conformation. **d** Overlay of the prepore and pore heptamers, viewed from the "top". Both the D1′ and D3 domains are essentially invariant, as are the first hairpin and helix of D2, in contrast to the major changes observed in the barrel of D2

and thereby ensure that unladen PA oligomers are not (wastefully) endocytosed. The ternary intermediate, which should be favored by high concentrations of toxic enzyme, also provides a satisfactory rationale for the existence of the octameric species (via the coalescence of four $(PA_{\mathbf{63}})_2$LF intermediates).

Structures of Toxin-Laden PA

In the crystal structure of $(PA_{\mathbf{63}})_8(LF\mathbf{N})_4$ (Feld et al. 2010), LF**N** perches on the outer rim of the heptamer formed by the exposed β-strands of PA D1′, and does indeed straddle two $PA_{\mathbf{63}}$ protomers. The major surprise was a major refolding of the N-terminal β-strand and helix (H1) of LF**N** to create a new interface (in the counterclockwise protomer). H1 sits in a new site (the "α-clamp"), promoting a "translocation-primed" conformation (Fig. 18.2a), which involves a movement of

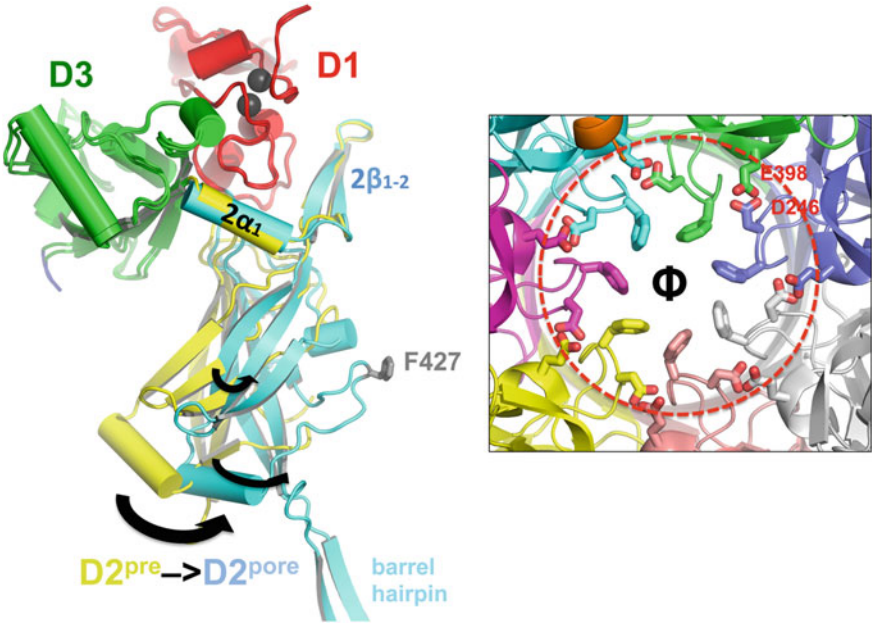


Fig. 18.5 A more detailed view of conformation changes on pore formation. At left, a comparison of the PA_{63} protomers from the pore and prepore, overlaid on the invariant D1′ and D3 domains. The pro-β-barrel segments have been removed to reveal the inward rotation of the D2 domain upon pore (shown in cyan) formation, as well as the stabilization of the loop carrying F427, seven of which coalesce to the form the Φ-clamp (shown at right), which is surrounded by a ring of 14 acidic residues, as shown. The view looks down the axis of the central β-barrel

~40 Å of helix 1α**1** of LF**N**. Upstream of 1α**1** are a further 30 residues that are disordered in the structure.

The structure of the PA_{63} heptamer bound to three full-length LF molecules (Fabre et al. 2016) confirmed the binding mode to the heptamer, which is mediated entirely by the N-terminal domains. The three LF molecules bind in the conformation observed in the free molecules; i.e., they dock essentially as essentially as rigid bodies. Intriguingly, the C-terminal tails bind to an N-terminal domain of their clockwise neighbors, most likely stabilizing the overall assembly, with the C-terminal domains suspended high (30–40 Å) above the heptamer. The odd number of protomers results in a single "orphan" PA_{63} that makes no contacts with LF, so that the arrangement creates a large head-to-tail unclosed "horseshoe", with three large holes in the structure (Fig. 18.3).

More recently, structures of full-length EF (which utilizes a homologous domain, EF**N**) bound to the pore have been determined by cryoEM at medium resolution (Zhou et al. 2020). These structures confirm that EF**N** binds in an identical fashion as LF**N** binds to the prepore. This was expected, but also confirms that the prepore to pore transition, per se, does not trigger a change in binding of the toxic enzymes. One surprise was that the C-terminal domains of EF are oriented in the opposite

(counterclockwise) direction to those of LF, but still make stabilizing contacts with the neighboring EF**N** domain (but on the opposite face). The structures are consistent with a maximum of 3 molecules of EF bound to heptameric pore/prepore,[1] as well as mixed EF/LF complexes (Pimental et al. 2004).

A Potential Role for PA_{83}/PA_{21} in Toxin Assembly

There is tantalizing evidence for a role for PA_{83} (and thus PA_{21}), either nicked or intact, in toxin assembly and prepore stability in some instances. For example, one report provides evidence for "mixed" heptamers containing both PA_{63} and PA_{83} that form functional pores in vivo when furin concentrations are limiting (Chekanov et al. 2006), and another report indicates that $PA_{\mathbf{21}}$ binds directly to LF in vitro (Chvyrkova et al. 2007). Indeed, simple in silico rigid-body modeling suggests that, while a single cleavage event must occur in PA to create an LF binding site, there are no steric impediments to binding of PA_{83} on either the clockwise or counterclockwise sides of the $PA_{\mathbf{63}}$:LF entity. In the latter case (which would create a ternary PA**63**:LF:PA**83** moiety similar to the PA**63**:LF:PA_{63} assembly intermediate described above), modeling predicts that the PA_{21} moiety of PA**83** would fit snugly against the N-terminal domain of LF bound to PA**63**, as well as providing support for LF's C-terminal domains (and plugging the holes in the heptameric complex). However, binding can occur only to the native conformation of LF, since PA**21** occupies the space that includes the α-clamp (Fig. 18.2b). Thus, PA_{83} (either intact or nicked) could provide additional stabilization for the ternary intermediate as well as the final oligomer, at the same time inhibiting the first step of LF/EF activation (and thus the timing of translocation) in the context of the fully-laden PA heptamer. Subsequent proteolytic/ejection events would be necessary to activate the toxin-laden pore. While speculative, these issues should be readily testable.

The Prepore-Pore Transition

The general nature of this remarkable conformational switch was predicted by Dr. Carlo Petosa more than 20 years ago (Petosa et al. 1997), and was subsequently supported by mutational, biophysical and, ultimately, a high resolution structure of the pore (Jiang et al. 2015).

The structure of the PA**63** heptameric pore was determined by single particle cryoEM methods in the presence of lipids at a resolution of 3.5 Å, sufficient for a complete trace of the polypeptide chain (Jiang et al. 2015) (Fig. 18.4c). Remarkably,

[1] All of the complexes described here are consistent with each other and with a large body of biochemical and biophysical studies; but they bear no resemblance to earlier EM-derived models of a PA-LF complex (Ren et al. 2004; Tama et al. 2006), which cannot be rationalized in terms of any of the other published complexes, or inferred from translocation intermediates, and so remains enigmatic.

the structure revealed that conformational changes are almost entirely restricted to D2 (although D4, once freed from its shackles to D2, becomes quite mobile, but still retains its MIDAS bond to the receptor). In contrast, D1′ and D3 remain essentially static: thus, an overlay of all Cα atoms within D1′ and D3 (as well as the connecting elements of D2, the N-terminal hairpin-helix segment) of the prepore and pore gives an RMS difference (for the entire heptameric annulus) of 0.68 Å, well within experimental error (Fig. 18.4d). In fact, this D1′-D3 annulus maintains the integrity of its docking platforms for the toxic enzymes throughout the translocation process.

The residues destined to form the membrane-penetrating portion of the central β-barrel comprise the mobile protease sensitive loop (residues 298–327). However, they lie at the top (*i.e.* membrane-distal) of the two flanking strands (2β2 and 2β3) which will form the upper part of the central β-barrel. These two strands must therefore first must first peel away from their parent domain, breaking countless intradomain H-bonds in the process. And in order for this to happen, D2 must first break its bonds with D4 and with the receptor.

What triggers these large reorganizations remains a matter for debate (*see* below). But what occurs is clear. The 2β2-2β3 hairpin lies at one end of the D2 barrel (distal to the heptamer axis. Its loss has little effect on the remainder of the D2′ barrel, which rotates essentially as a rigid body ~30° "inward" toward the heptamer axis. Only the final β-hairpin and helix of D2, which pack against D1′, remain static during this process. The rotation has two important effects.

First, the rotation brings the ends of strands 2β**1** and 2β**4** (which "hold" the ends of new β-hairpin) ~12 Å closer to the heptamer axis (and ~20 Å closer to their symmetry-related counterparts). This closure enables the 7 long β-hairpins to associate into a 14-stranded β-barrel ~100 Å in length, similar in nature to that found in hemolysin, but roughly twice as long (Fig. 18.4c).

Second, the rotation brings the seven D2′ barrels into close apposition, creating a compact interface that forms the remaining elements of the protein translocation machine within the body of the heptamer. In particular, seven Phe427 residues are brought together to form a flexible annulus with a small hole at its center; this is called the "Φ-clamp" and its integrity is absolutely key to successful protein translocation. It is complemented with rings of acidic residues that surround the Φ-clamp on both the endosomal and cytoplasmic sides (Figs. 18.4 and 18.5).

Protein Translocation

An extensive series of studies, mostly by Krantz, Collier, Leppla and colleagues (Wynia-Smith et al. 2012; Krantz et al. 2005, 2006), has elucidated the basic elements of translocation of the toxic enzymes through the PA pore from endosome to cytosol. Briefly, each ~90 kDa enzyme must unfold, starting at its N-terminus (Gupta et al. 2008), and traverse the Φ-clamp and β-barrel as an extended polypeptide, one at a time; and then refold once in the cytosol. Binding of the toxic enzymes to the rim of the heptamer primes them for this process, and their N-terminal rearrangement

creates a platform for the N-terminal helix (the α-clamp) of LF**N**/EF**N** to bind at a suitable height for its disordered ~30-residue N-terminus to hang down and make first contact with the Φ-clamp, ~45 Å below the α-clamp (Brown et al. 2015).

The Φ-clamp is watertight (when peptide is engaged), but flexible enough to allow positively charged and large aromatic residues (both of which readily interact with the ring of Phe residues via hydrophobic and/or Π-cation forces) to pass through the pore in either direction; while negatively charged residues cannot. However, the pH of the late endosome is just low enough to protonate Asp/Glu (pKa = 4.0–4.4), at least transiently, enabling them to pass through the Φ-clamp and onto the cytoplasmic side, where the much higher pH immediately deprotonates them, effectively trapping them on the distal side. That is, the pore provides a conduit for a single extended polypeptide to traverse, but it is protonation/deprotonation of Asp/Glu residues that is the key process for driving protein translocation from endosome to cytosol (by defining irreversible directionality). The mechanism is typically referred to as a "Brownian ratchet"—imagined as random oscillations of peptide segments lacking acidic residues (back and forth across the clamp), punctuated by periodic, irreversible, unidirectional translocational steps, whenever protonated Asp/Glu are encountered within the polypeptide chain; these have been called the "teeth" of the ratchet.

Trigger for the Prepore to Pore Transition

Finally, what drives the conformational change from pre-pore to pore? We know that, in vitro, $PA_{\mathbf{63}}$ heptamers insert into synthetic bilayers (in the absence of receptor) to form functional pores when the pH is lowered to ~6.0, which strongly implicates the protonation of histidine residues as triggers for the change. The lower pH required for membrane insertion in the presence of the receptor, CMG2 (pH ~ 5) (Scobie et al. 2007) is consistent with additional restrictions imposed by the very strong interactions (Kd ~170 pM) between PA and CMG2 that lock D2 and D4 together, thereby lowering the pKa values of certain histidines. The evidence that pore formation occurs in early endosomal compartments, while protein translocation occurs later, at a lower pH, would also argue against a role for Asp/Glu titration.

Exhaustive point mutational studies of PA_{63} showed that the substitution of any single histidine (by cysteine) did not impede pore formation, suggesting that protonation of more than one histidine is necessary. Six of the nine histidines in PA**63**, do not appear to be relevant to the pH switch: thus, His304, His310, His253 and His597 have weak or negligible interactions in the prepore and are unchanged or solvent-exposed in the pore; while His616 is buried in the center of D4, which remains folded following pore formation; and His211 is on the (invariant) surface of D1′ at the top of the heptamer (where it interacts with LF/EF in a pH-independent manner).

However, there are three histidines (His263, His299 and His336) PA_{63} that change from buried to an exposed environment upon switching from prepore to pore; moreover, their locations at key points within D2 or at the D2-D4 interface make them plausible candidates for contributing to the pH trigger [Fig. 18.6]. In addition, a

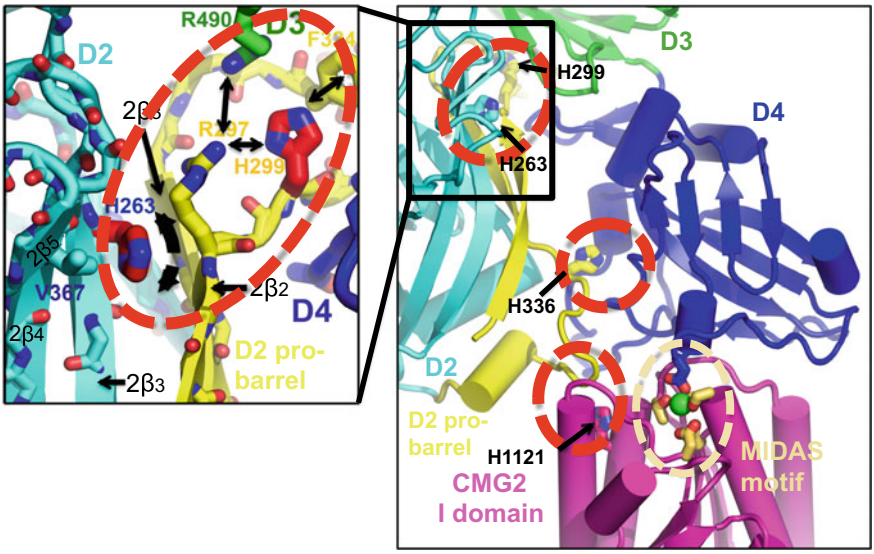


Fig. 18.6 Potential pH triggers in the context of the PA-CMG2 complex. At right is an overview of the receptor complex, showing the MIDAS-mediated interaction with PA D4, and the interaction with a pro-barrel region of D2, which locks PA in its prepore state. Four histidines are identified that might contribute to the pH trigger: H299, H263 and H336 from D2, and H1121 from CMG2. At left is an expanded view of the environment of two of these histidines. Note how H263 and H299 are held in place by hydrophobic interactions but also form a sandwich around an Arg residue from D3. In addition, H263 inserts into the top of the b-sheet of D4, disrupting interactions that would create a barrel. Protonation of any of these residues is predicted to destabilize an interface, and thereby promote unfolding of the D4 domain

conserved residue, His1121, from CMG2, is packed against Arg 659 (D4) and the D2 loop in the receptor complex, and is exposed to solvent following transition to the pore, such that its protonation could also promote the switch.

Krantz and colleagues have argued that residues close to the Φ-clamp are somehow involved in the pH trigger, including one Asp and 2 non-titratable residues (Jiang et al. 2015). Their rationale seems to be that these residues form dominant negative mutants, but it is unclear what this has to do with pH-dependent triggers. Notwithstanding, it is clear that identifying such triggers is difficult, and does not readily lend itself to simple mutational approaches.

References

Abrami L, Lindsay M, Parton RG, Leppla SH, van der Goot FG (2004) Membrane insertion of anthrax protective antigen and cytoplasmic delivery of lethal factor occur at different stages of the endocytic pathway. J Cell Biol 166(5):645–651. https://doi.org/10.1083/jcb.200312072

Abrami L, Liu S, Cosson P, Leppla SH, van der Goot FG (2003) Anthrax toxin triggers endocytosis of its receptor via a lipid raft-mediated clathrin-dependent process. J Cell Biol 160(3):321–328. https://doi.org/10.1083/jcb.200211018

Bradley KA, Mogridge J, Mourez M, Collier RJ, Young JA (2001) Identification of the cellular receptor for anthrax toxin. Nature 414(6860):225–229. https://doi.org/10.1038/n35101999/n35101999[pii]

Brown MJ, Thoren KL, Krantz BA (2015) Role of the alpha clamp in the protein translocation mechanism of anthrax toxin. J Mol Biol 427(20):3340–3349. https://doi.org/10.1016/j.jmb.2015.08.024

Chekanov AV, Remacle AG, Golubkov VS, Akatov VS, Sikora S, Savinov AY, Fugere M, Day R, Rozanov DV, Strongin AY (2006) Both PA63 and PA83 are endocytosed within an anthrax protective antigen mixed heptamer: a putative mechanism to overcome a furin deficiency. Arch Biochem Biophys 446(1):52–59. https://doi.org/10.1016/j.abb.2005.11.013

Chvyrkova I, Zhang XC, Terzyan S (2007) Lethal factor of anthrax toxin binds monomeric form of protective antigen. Biochem Biophys Res Commun 360(3):690–695. https://doi.org/10.1016/j.bbrc.2007.06.124

Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC (2000) Structural basis of collagen recognition by integrin alpha2beta1. Cell 101(1):47–56. https://doi.org/10.1016/S0092-8674(00)80622-4

Fabre L, Santelli E, Mountassif D, Donoghue A, Biswas A, Blunck R, Hanein D, Volkmann N, Liddington R, Rouiller I (2016) Structure of anthrax lethal toxin prepore complex suggests a pathway for efficient cell entry. J General Physiol 148(4):313–324. https://doi.org/10.1085/jgp.201611617

Feld GK, Thoren KL, Kintzer AF, Sterling HJ, Tang II, Greenberg SG, Williams ER, Krantz BA (2010) Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. Nat Struct Mol Biol 17(11):1383–1390. https://doi.org/10.1038/nsmb.1923

Frankel AE, Kuo SR, Dostal D, Watson L, Duesbery NS, Cheng CP, Cheng HJ, Leppla SH (2009) Pathophysiology of anthrax. Front Biosci 14:4516–4524. 3544 [pii]

Gordon VM, Klimpel KR, Arora N, Henderson MA, Leppla SH (1995) Proteolytic activation of bacterial toxins by eukaryotic cells is performed by Furin and by additional cellular proteases. Infect Immun 63(1):82–87

Gu ML, Leppla SH, Klinman DM (1999) Protection against anthrax toxin by vaccination with a DNA plasmid encoding anthrax protective antigen. Vaccine 17 (4):340–344. S0264-410X(98)00210-2 [pii]

Gupta PK, Moayeri M, Crown D, Fattah RJ, Leppla SH (2008) Role of N-terminal amino acids in the potency of anthrax lethal factor. PLoS ONE 3(9):e3130. https://doi.org/10.1371/journal.pone.0003130

Jiang J, Pentelute BL, Collier RJ, Zhou ZH (2015) Atomic structure of anthrax protective antigen pore elucidates toxin translocation. Nature 521(7553):545–549. https://doi.org/10.1038/nature14247

Kintzer AF, Sterling HJ, Tang II, Williams ER, Krantz BA (2010) Anthrax toxin receptor drives protective antigen oligomerization and stabilizes the heptameric and octameric oligomer by a similar mechanism. PLoS ONE 5(11):e13888. https://doi.org/10.1371/journal.pone.0013888

Kintzer AF, Thoren KL, Sterling HJ, Dong KC, Feld GK, Tang II, Zhang TT, Williams ER, Berger JM, Krantz BA (2009) The protective antigen component of anthrax toxin forms functional octameric complexes. J Mol Biol 392(3):614–629. https://doi.org/10.1016/j.jmb.2009.07.037

Krantz BA, Finkelstein A, Collier RJ (2006) Protein translocation through the anthrax toxin transmembrane pore is driven by a proton gradient. J Mol Biol 355(5):968–979. https://doi.org/10.1016/j.jmb.2005.11.030

Krantz BA, Melnyk RA, Zhang S, Juris SJ, Lacy DB, Wu Z, Finkelstein A, Collier RJ (2005) A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. Science 309(5735):777–781. https://doi.org/10.1126/science.1113380

Lacy DB, Wigelsworth DJ, Melnyk RA, Harrison SC, Collier RJ (2004) Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation. Proc Natl Acad Sci USA 101(36):13147–13151. https://doi.org/10.1073/pnas.0405405101

Larson DK, Calton GJ, Little SF, Leppla SH, Burnett JW (1988) Separation of three exotoxic factors of Bacillus anthracis by sequential immunosorbent chromatography. Toxicon 26(10):913–921

Leppla SH, Robbins JB, Schneerson R, Shiloach J (2002) Development of an improved vaccine for anthrax. J Clin Invest 110(2):141–144. https://doi.org/10.1172/JCI16204

Moayeri M, Leppla SH (2004) The roles of anthrax toxin in pathogenesis. Curr Opin Microbiol 7(1):19–24. https://doi.org/10.1016/j.mib.2003.12.001/S1369527403001668[pii]

Moayeri M, Leppla SH (2009) Cellular and systemic effects of anthrax lethal toxin and edema toxin. Molecular Aspects Med 30 (6):439–455. S0098-2997(09)00051-X [pii]/10.1016/j.mam.2009.07.003

Moayeri M, Leppla SH, Vrentas C, Pomerantsev AP, Liu S (2015) Anthrax pathogenesis. Annu Rev Microbiol 69:185–208. https://doi.org/10.1146/annurev-micro-091014-104523

Mogridge J, Cunningham K, Lacy DB, Mourez M, Collier RJ (2002) The lethal and edema factors of anthrax toxin bind only to oligomeric forms of the protective antigen. Proc Natl Acad Sci USA 99(10):7045–7048. https://doi.org/10.1073/pnas.052160199

Molloy SS, Bresnahan PA, Leppla SH, Klimpel KR, Thomas G (1992) Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. J Biol Chem 267(23):16396–16402

Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC (1997) Crystal structure of the anthrax toxin protective antigen. Nature 385(6619):833–838. https://doi.org/10.1038/385833a0

Pilpa RM, Bayrhuber M, Marlett JM, Riek R, Young JA (2011) A receptor-based switch that regulates anthrax toxin pore formation. PLoS Pathog 7(12):e1002354. https://doi.org/10.1371/journal.ppat.1002354

Pimental RA, Christensen KA, Krantz BA, Collier RJ (2004) Anthrax toxin complexes: heptameric protective antigen can bind lethal factor and edema factor simultaneously. Biochem Biophys Res Commun 322(1):258–262. https://doi.org/10.1016/j.bbrc.2004.07.105

Ren G, Quispe J, Leppla SH, Mitra AK (2004) Large-scale structural changes accompany binding of lethal factor to anthrax protective antigen: a cryo-electron microscopic study. Structure 12(11):2059–2066. https://doi.org/10.1016/j.str.2004.09.010

Rosovitz MJ, Schuck P, Varughese M, Chopra AP, Mehra V, Singh Y, McGinnis LM, Leppla SH (2003) Alanine-scanning mutations in domain 4 of anthrax toxin protective antigen reveal residues important for binding to the cellular receptor and to a neutralizing monoclonal antibody. J Biol Chem 278(33):30936–30944. https://doi.org/10.1074/jbc.M301154200/M301154200[pii]

Santelli E, Bankston LA, Leppla SH, Liddington RC (2004) Crystal structure of a complex between anthrax toxin and its host cell receptor. Nature 430(7002):905–908. https://doi.org/10.1038/nature02763

Scobie HM, Marlett JM, Rainey GJ, Lacy DB, Collier RJ, Young JA (2007) Anthrax toxin receptor 2 determinants that dictate the pH threshold of toxin pore formation. PLoS ONE 2(3):e329. https://doi.org/10.1371/journal.pone.0000329

Singh Y, Klimpel KR, Arora N, Sharma M, Leppla SH (1994) The chymotrypsin-sensitive site, FFD315, in anthrax toxin protective antigen is required for translocation of lethal factor. J Biol Chem 269(46):29039–29046

Singh Y, Klimpel KR, Goel S, Swain PK, Leppla SH (1999) Oligomerization of anthrax toxin protective antigen and binding of lethal factor during endocytic uptake into mammalian cells. Infect Immun 67(4):1853–1859

Tama F, Ren G, Brooks CL 3rd, Mitra AK (2006) Model of the toxic complex of anthrax: responsive conformational changes in both the lethal factor and the protective antigen heptamer. Protein Sci 15(9):2190–2200. https://doi.org/10.1110/ps.062293906

van der Goot G, Young JA (2009) Receptors of anthrax toxin and cell entry. Mol Aspects Med 30(6):406–412. https://doi.org/10.1016/j.mam.2009.08.007

Varughese M, Teixeira AV, Liu S, Leppla SH (1999) Identification of a receptor-binding region within domain 4 of the protective antigen component of anthrax toxin. Infect Immun 67(4):1860–1865

Wynia-Smith SL, Brown MJ, Chirichella G, Kemalyan G, Krantz BA (2012) Electrostatic ratchet in the protective antigen channel promotes anthrax toxin translocation. J Biol Chem 287(52):43753–43764. https://doi.org/10.1074/jbc.M112.419598

Young JA, Collier RJ (2007) Anthrax toxin: receptor binding, internalization, pore formation, and translocation. Annu Rev Biochem 76:243–265. https://doi.org/10.1146/annurev.biochem.75.103004.142728

Zhou K, Liu S, Hardenbrook NJ, Cui Y, Krantz BA, Zhou ZH (2020) Atomic structures of anthrax prechannel bound with full-length lethal and edema factors. Structure. https://doi.org/10.1016/j.str.2020.05.009