David A.D. Parry
John M. Squire *Editors*

Fibrous Proteins: Structures and Mechanisms

Springer

Subcellular Biochemistry

Volume 82

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.

Series Editor
J. Robin Harris, University of Mainz, Mainz, Germany

International Advisory Editorial Board
T. Balla, National Institutes of Health, NICHD, Bethesda, USA
Tapas K. Kundu, JNCASR, Bangalore, India
A. Holzenburg, Texas A&M University, College Station, USA
S. Rottem, The Hebrew University, Jerusalem, Israel
X. Wang, Jiangnan University, Wuxi, China

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

David A.D. Parry • John M. Squire

Editors

Fibrous Proteins: Structures and Mechanisms

Springer

Editors
David A.D. Parry
Institute of Fundamental Sciences
    and Riddet Institute
Massey University
Palmerston North, New Zealand

John M. Squire
Muscle Contraction Group, School
    of Physiology, Pharmacology
    and Neuroscience
University of Bristol
Bristol, UK

Preface

The first insight into the secondary structures of proteins came in the 1950s from fibre X-ray diffraction data and model-building studies based on stereochemical constraints and hydrogen-bonding potential. These discoveries, and later work on the three-dimensional structure of globular proteins using protein crystallography and the development of protein sequencing methods, led to a flurry of Nobel Prizes, including those to Linus Pauling, Max Perutz, John Kendrew and Fred Sanger. Apart from water, fibrous proteins constitute the bulk of our bodies and are integral in defining our structural form and giving us the ability to move (amongst other things). Fibrous proteins are present intracellularly as intermediate filaments and myosin and actin filaments and extracellularly as collagen fibrils. They are particularly abundant as collagen in bone, tendon and cartilage and as parts of the myosin and actin filaments in muscle, including the heart. Intermediate filaments are also abundant as keratin in skin, hair and nails. Historically, the regular nature of the amino acid sequences in fibrous proteins has made them more amenable to structural analysis than has been the case for their globular cousins. This has had the benefit that their study has led to insights into a number of diseases. Importantly, research into fibrous protein structures, along with the structures of synthetic analogues, has led to rules about how protein chains fold into such conformations as α-helical coiled coils, β-sheets or collagen triple helices. Indeed, sequences can now be designed to generate structures with defined functional properties. This, in turn, has led to the relatively new realm of protein engineering. Novel methods of generating and processing fibrous proteins, such as silks, connective tissues and coiled-coil proteins, have also led to the preparation of twenty-first-century biomaterials with properties that are both useful and of significant potential in industry and medicine.

The fibrous protein field was last surveyed in detail in 2005/2006, when a series of books was published as part of the Advances in Protein Chemistry series (Vol. 70, *Coiled-coils, Collagen and Elastomers* (2005); Vol. 71, *Fibrous Proteins: Muscle and Molecular Motors* (2005); Vol. 73, *Fibrous Proteins: Amyloids, Prions and Beta Proteins* (2006)). Much has happened in the field over the intervening years, so the present volume *Fibrous Proteins: Structures and Mechanisms* aims to bring

coverage of the field up to date. The chapters have all been written by world experts in their own particular field and are a 'state-of the-art' summary of what is now known. *Fibrous Proteins: Structures and Mechanisms* will be a valuable resource for those working in the field, both for senior scientists and for new graduate and postgraduate students. It starts with a historical overview of the field, to bring new-comers up to speed. The book as a whole is clearly written and liberally illustrated and not only describes what is known but also discusses the applications of protein engineering and the commercial exploitation of new biomaterials. Furthermore, it establishes the basis for deciding the most appropriate directions for future research.

Those interested in being a continuing part of the fibrous protein field may wish to attend the Workshops on Fibrous Proteins held at Alpbach, Austria, every 4 years. Apart from being in a delightful Alpine setting, these Workshops bring together scientists from across the world at the cutting edge of modern fibrous protein research. Details of the Workshops, started and organised by the current editors from 1993 to 2009 and now run by Andrei Lupas (see Chap. 4) and Dek Woolfson (see Chap. 2), can be found on the Workshop website at http://www.coiledcoils.org/.

Palmerston North, New Zealand   David A.D. Parry
Bristol, UK   John M. Squire

Contents

Chapter 1
Fibrous Protein Structures: Hierarchy, History and Heroes

John M. Squire and David A.D. Parry

Contents

Abstract During the 1930s and 1940s the technique of X-ray diffraction was applied widely by William Astbury and his colleagues to a number of naturally-occurring fibrous materials. On the basis of the diffraction patterns obtained, he observed that the structure of each of the fibres was dominated by one of a small number of different types of molecular conformation. One group of fibres, known as the k-m-e-f group of proteins (keratin – myosin – epidermin – fibrinogen), gave rise to diffraction characteristics that became known as the α-pattern. Others, such

J.M. Squire (✉)
Muscle Contraction Group, School of Physiology, Pharmacology and Neuroscience,
University of Bristol, Bristol BS8 1TD, UK
e-mail: j.squire@imperial.ac.uk

D.A.D. Parry
Institute of Fundamental Sciences and Riddet Institute, Massey University,
Private Bag 11-222, Palmerston North 4442, New Zealand

as those from a number of silks, gave rise to a different pattern – the β-pattern, while connective tissues yielded a third unique set of diffraction characteristics. At the time of Astbury's work, the structures of these materials were unknown, though the spacings of the main X-ray reflections gave an idea of the axial repeats and the lateral packing distances. In a breakthrough in the early 1950s, the basic structures of all of these fibrous proteins were determined. It was found that the long protein chains, composed of strings of amino acids, could be folded up in a systematic manner to generate a limited number of structures that were consistent with the X-ray data. The most important of these were known as the α-helix, the β-sheet, and the collagen triple helix. These studies provided information about the basic building blocks of all proteins, both fibrous and globular. They did not, however, provide detailed information about how these molecules packed together in three-dimensions to generate the fibres found *in vivo*. A number of possible packing arrangements were subsequently deduced from the X-ray diffraction and other data, but it is only in the last few years, through the continued improvements of electron microscopy, that the packing details within some fibrous proteins can now be seen directly. Here we outline briefly some of the milestones in fibrous protein structure determination, the role of the amino acid sequences and how new techniques, including electron microscopy, are helping to define fibrous protein structures in three-dimensions. We also introduce the idea that, from the known sequence characteristics of different fibrous proteins, new molecules can be designed and synthesized, thereby generating new biological materials with specific structural properties. Some of these, for example, are planned for use in drug delivery systems. Along the way we also introduce the various Chapters of the book, where individual fibrous proteins are discussed in detail.

Keywords α-helix • Coiled-coil • β-pleated sheet • Cross-β structure • Collagen fold • Heptad • Stutters • Stammers • Amino acid sequence

List of Abbreviations

TEM	transmission electron microscopy
SEM	scanning electron microscopy
STEM	scanning-transmission electron microscopy
PBLG	poly-γ-benzyl-L-glutamate
DNA	deoxyribonucleic acid
NMR	nuclear magnetic resonance
D	67 nm period in collagen fibrils
CCD	charged couple device
CMOS	complementary metal-oxide semiconductor
ELT	enclosed-gate layout transistors
MTF	modulation transfer function

1.1   Introduction: Early History and Key Players

1.1.1   Simplicity and Complexity in Amino Acids

Biology today is dominated by our ability to solve the structures of proteins, poly-nucleotides, polysaccharides, viruses and other molecules of biological importance at atomic resolution, thereby revealing the mechanisms by which they function as integral members of some physiological system. This capability, however, arose only at the beginning of the second half of the twentieth century, and was initiated by research into the structure of proteins. Proteins were then known to be polypep-tides, i.e. strings of amino acids covalently linked through what became known as peptide bonds. The chemistry of amino acids is deceptively simple, as each amino acid has a common formula NH_2-C_αHR-C'OOH. The core of the amino acid is the alpha carbon (C_α), which is a tetrahedral carbon atom making single bond interac-tions with four other atoms or groups. These are:

1. A trigonal carbon atom (C′) linked to an OH (a hydroxyl group) and double-bonded to an oxygen atom (a carbonyl group)
2. A trigonal nitrogen with two hydrogen atoms (an amine group)
3. A lone hydrogen atom
4. An R-group, which represents the amino acid sidechain. The R-groups are char-acteristic of each of the 20 different amino acids, as detailed later.

Two amino acids can link, as in Fig. 1.1, through a condensation process which results in the formation of a peptide bond and the release of a water molecule. This arises when the NH_2 (amine) group of one amino acid loses a hydrogen atom and the C′OOH (carboxyl) group of a second amino acid loses an OH. The resulting bond formed between the NH group of amino acid 1 and the C′O group of amino acid 2 is known as the peptide bond. When many amino acids are linked in this way, they form a polypeptide chain. All proteins have this same basic structure, the dif-ferences between them being determined solely by the particular sequences of R groups that occur along the chains.

The first real evidence of regularities in protein folding came from the pioneering X-ray diffraction studies of a variety of protein fibres by William Astbury (Fig. 1.2a; see, for example, Astbury and Woods 1930, 1933; Astbury and Dickinson 1940; Astbury and Bell 1941; Astbury 1949a, b, 1951), who studied hair, muscle, connec-tive tissue, silks and many other tissues. He found that he could classify some of these materials as having similar X-ray diffraction patterns. For example, one group, known as the k-m-e-f group of proteins, gave a particular type of diffraction pattern with a strong diffraction peak in the same direction as the fibre axis (the meridian) at a spacing of about 0.51 nm and a maximum on the equator with a spacing of about 1 nm. This was called the α-pattern. The letters k-m-e-f refer to keratin (in hair, wool, claw, hoof, horn, quill and baleen), myosin (in muscle), epidermin (in skin) and fibrinogen (in blood clots). Others, such as *Bombyx mori* silk, gave a different diffraction pattern (the β–pattern) characterized by maxima on the meridian
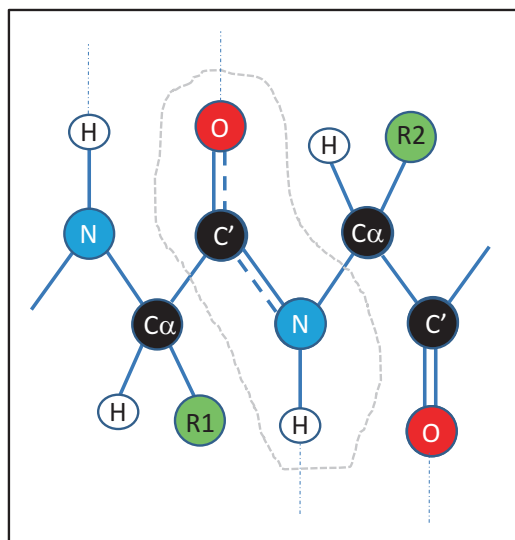
Fig. 1.1 Schematic diagram of two amino acids ($NH_2.C_\alpha HR.C'OOH$) linked by a peptide bond. The $C'OOH$ group of one amino acid and the NH_2 group of its successor bond through a condensation reaction where the hydroxyl group from the C' atom and a hydrogen from the amine group are eliminated as water. The central OC'-NH group (*dotted outline*) is the amide group. The NH and CO groups have the potential to make hydrogen bonds with other amino acids, as illustrated by the *vertical dashed lines*. Different amino acids are characterized by their unique R groups on the C_α, illustrated later as Fig. 1.7

and equator with spacings of about 0.35 and 0.48 nm, respectively. Connective tissues gave yet a third distinct set of diffraction maxima (the collagen pattern) with a strong meridional reflection at a spacing of 0.29 nm and pronounced equatorial reflections with spacings of about 1.2 and 0.6 nm.

It is of interest here that a comprehensive and very impressive study of the naturally-occurring silks, undertaken by Rudall and colleagues (see Fig. 1.2 caption), showed that while most of these silks exhibited a β-pattern, there were examples of both the α- and the collagen pattern too. For example, the silk from bees, wasps and ants (*Hymenoptera aculeata*) gave an α-pattern (Rudall 1962, 1965; Atkins 1967; Lucas and Rudall 1968), whilst the silk of the gooseberry sawfly (*Nematus ribesii*) gave a collagen pattern (Lucas and Rudall 1968).

In the late 1940s and early 1950s biological scientists started to think seriously about the manner in which the polypeptide chain might fold up in three-dimensions to give the observed diffraction patterns. A crucial insight came from the realization that a $C'=O$ group from one amino acid residue and an N-H group of a different amino acid might interact with each other through the formation of a hydrogen bond, and do so in a systematic and long-range manner. Available at that time were bond angles and bond lengths that had come from earlier X-ray diffraction studies of crystals of small molecules (Corey and Donohue 1950). On this basis, Lawrence Bragg, John Kendrew and Max Perutz (Bragg et al. 1950), all Nobel Laureates, or

soon to become so (Fig. 1.2b, c, d), and all at the Cavendish Laboratory in Cambridge, UK, carried out a systematic study of possible hydrogen-bonding schemes, where the hydrogen-bonded N-H and C′=O groups were present in the same polypeptide chain. Following on from the seminal work of Huggins (1943) and Crane (1950), who had suggested that any regular conformation of a protein chain would be such that the main chain atoms of every residue would be structurally equivalent (and hence in a helical conformation), Bragg, Kendrew and Perutz came up with a number of possible structures. Two of these were a three-residue per turn (3_1) helix, and a four-residue per turn (4_1) helix (the ω-helix). Unfortunately, however, they only considered helices with an integral number of amino acids per turn. In addition, they also missed an important point about the nature of the peptide link. Linus Pauling (another Nobel Laureate), Robert Corey and Herman Branson (Pauling et al. 1951; Fig. 1.2e, f) had realized that the trigonal carbon, the trigonal nitrogen and the oxygen in the HN-C′O group (the amide group) would all have unpaired electrons in p orbitals, mainly above and below the plane of the trigonally-arranged single bonds. These electron p-orbitals would coalesce to give molecular orbitals (electron streamers or π bonds) concentrated above and below the plane of the trigonal bonds and would tend to make the amide group flat or planar. Looking at this another way, the π bonds would give the N-C′ and C′=O bonds partial double bond character (as in Fig. 1.1) and thus restrict rotation around the axes of these bonds. The amide groups would therefore tend to be planar, and would provide a strong constraint on possible structures stabilized by regular patterns of hydrogen bonding.

With this amide group structure in mind, and with the realisation that there was no particular reason why there should be an integral number of amino acids per turn of the helix, Pauling et al. (1951) explored various hydrogen-bonding schemes. They found a structure that would more or less explain the observed 0.51 nm meridional reflection in X-ray patterns from the α-proteins, that would keep the amide group planar, and that would have the satisfactory property of hydrogen bonds lying almost parallel to the axis of the helix, as previously inferred from infra-red spectroscopy by Elliott and Ambrose (1950). They called this structure the α-helix (Fig. 1.4) and used it to explain α–pattern structures (Pauling and Corey 1951a), as defined earlier by Astbury and colleagues.

1.1.2 The α-Helix

In its simplest form the α-helix has 18 amino acid residues in 5 turns of the helix (an 18_5 helix). The pitch of the helix would be about 0.54 nm, close to the 0.51 nm observed by Astbury, and the axial rise per residue would be expected to be about 0.15 nm. Having missed the α-helix in his earlier analysis with Bragg and Kendrew (Bragg et al. 1950), Max Perutz (Fig. 1.2e) considered how he might best test for the presence of the α-helix in known structures. From his understanding of the α-helical structure, allied to the helical diffraction theory that was being developed at about the same time (Cochran et al. 1952), Perutz predicted that there should be a strong

Fig. 1.2 Some of the heroes of the fibrous protein story. (**a**) William Thomas Astbury (1898–1961), English physicist and molecular biologist, who pioneered the study of fibrous materials by X-ray diffraction working mainly at the University of Leeds, UK. (**b**) Sir (William) Lawrence Bragg (1890–1971), Australian-born then British physicist and X-ray crystallographer, discoverer of Bragg's Law of Diffraction, joint winner with his father Sir William Bragg of the Nobel Prize for Physics in 1915, (**c**) John Cowdery Kendrew (1917–1997), English biochemist and crystallographer, who worked with Lawrence Bragg and Max Perutz at the Cavendish Laboratory (and subsequently the MRC Laboratory of Molecular Biology) in Cambridge, UK on possible ways that

Fig. 1.2 (continued) polypeptide chains might fold to produce helical structures. He went on to solve the structure of myoglobin. Kendrew shared the 1962 Nobel Prize for chemistry with Max Perutz for determining the first atomic structures of proteins using X-ray crystallography. (**d**) Max Ferdinand Perutz (1914–2002), Austrian-born and later British molecular biologist and crystallographer, who worked at the Cavendish Laboratory (and subsequently the MRC Laboratory of Molecular Biology) at Cambridge, UK. In 1962, he shared the Nobel Prize for chemistry with Kendrew (see (**c**) *above*). Perutz was part of the team, with Bragg and Kendrew, that suggested possible helically folded forms of the polypeptide chain. He then went on to prove the existence of the α–helix. (**e**) Linus Pauling (1901–1994), US chemist and molecular biologist, co-discoverer of the α-helix and β-pleated sheet, winner of two Nobel prizes (Chemistry in 1954 and Peace in 1962), working mainly at the California Institute of Technology, USA. He first made his name in 1939 with a book "The Nature of the Chemical Bond and the Structure of Molecules and Crystals" (Cornell University Press), which became a classic. (**f**) Robert Brainard Corey (1897–1971), US chemist at CalTech, who co-authored many papers with Pauling on the α–helix, β-sheet and the structures of fibrous proteins. (**g**) Arthur Elliott (1904–1996), UK physicist, molecular biologist and inventor, worked at the Courtaulds Laboratory in Maidenhead, UK and then at the Biophysics Department, Kings College, University of London, UK. Elliott and his colleagues at Courtaulds studied synthetic polypeptides, such as poly-γ-benzyl-L-glutamate (PBLG), and was able to supply Perutz with an oriented sample of PBLG with which Perutz proved the existence of the α–helix. He had previously shown by infra-red spectroscopy that the hydrogen-bonds in the helical forms of these polypeptides were aligned parallel to the helix axis. Elliott went on to study the packing of coiled coils in the paramyosin filaments in molluscan muscles. (**h**) Francis Harry Compton Crick (1916–2004), UK biophysicist and molecular biologist, working mainly at the Cavendish Laboratory (and subsequently the Laboratory of Molecular Biology) in Cambridge, UK is most famous for his discovery with Jim Watson and Maurice Wilkins of the double-helical structure of DNA, for which they were awarded the 1962 Nobel Prize in Physiology or Medicine. Before that, however, Crick made several major contributions to the fibrous protein field, particularly for his part in developing helical diffraction theory and for his proposal of the structure and expected diffraction pattern of the coiled-coil. (**i**) Fred Sanger (1918–2013), British biochemist working at the MRC Laboratory of Molecular Biology in Cambridge, UK who won the Nobel Prize in Chemistry twice, first for inventing a method to determine protein sequences (initially applied to insulin, Sanger 1959) and second for inventing a method of sequencing the bases in DNA. (**j**) Carolyn Cohen (b1926), US biologist, worked primarily at the Jimmy Fund Building in Boston, USA and at Brandeis University, Waltham, USA. Her major contributions lay in determining the crystal structures of tropomyosin, myosin and fibrinogen, and in her studies on the folding problem in proteins, with special regard to the importance of the heptad repeat in α-fibrous proteins. (**k**) Alexander Rich (1924–2015), US biologist and biophysicist, who worked at Massachusetts Institute of Technology and Harvard Medical School, is probably most well-known for his discovery of left–handed DNA. However, prior to that he worked with Francis Crick on the structure of collagen and in 1955 published a seminal paper outlining what proved to be the correct structure. They were aided in this determination by optical diffraction studies carried out by Arthur Elliott. (**l**) Robert Donald Bruce Fraser (b1924), Australian physicist and mathematician, who worked mainly at the Division of Protein Chemistry, CSIRO, Melbourne, Australia (often with his long-term colleague Tom MacRae). His earlier work was at Kings College, London, UK where, among other things, he proposed a structure for DNA. At that time he was working with Maurice Wilkins (Nobel Laureate) and Rosalind Franklin of DNA fame. He moved to the Division of Protein Chemistry, Melbourne, Australia in the early 1950s and became a leading light in the fibrous protein field, particularly in his studies of α–helices in intermediate filaments, β-structures in silks and keratins and molecular packing in collagen fibrils. His outstanding 1973 book with MacRae entitled "Conformation in Fibrous Proteins" (Academic Press, New York and London) became a central handbook in the field and remains very useful to this day. A final hero (not photographed here) is Kenneth Maclaurin Rudall (1910–1996), New Zealand biophysicist, worked at Leeds University, UK who was a supreme experimentalist. His ability to prepare highly oriented specimens suitable for fibre X-ray diffraction has never been equalled. His main contributions lay in the fields of chitin, epidermin and silk structure, and his studies were instrumental in recognising the cross-β conformation. He also studied structural transitions in the k-m-e-f group of α-fibrous proteins caused by extension and contraction
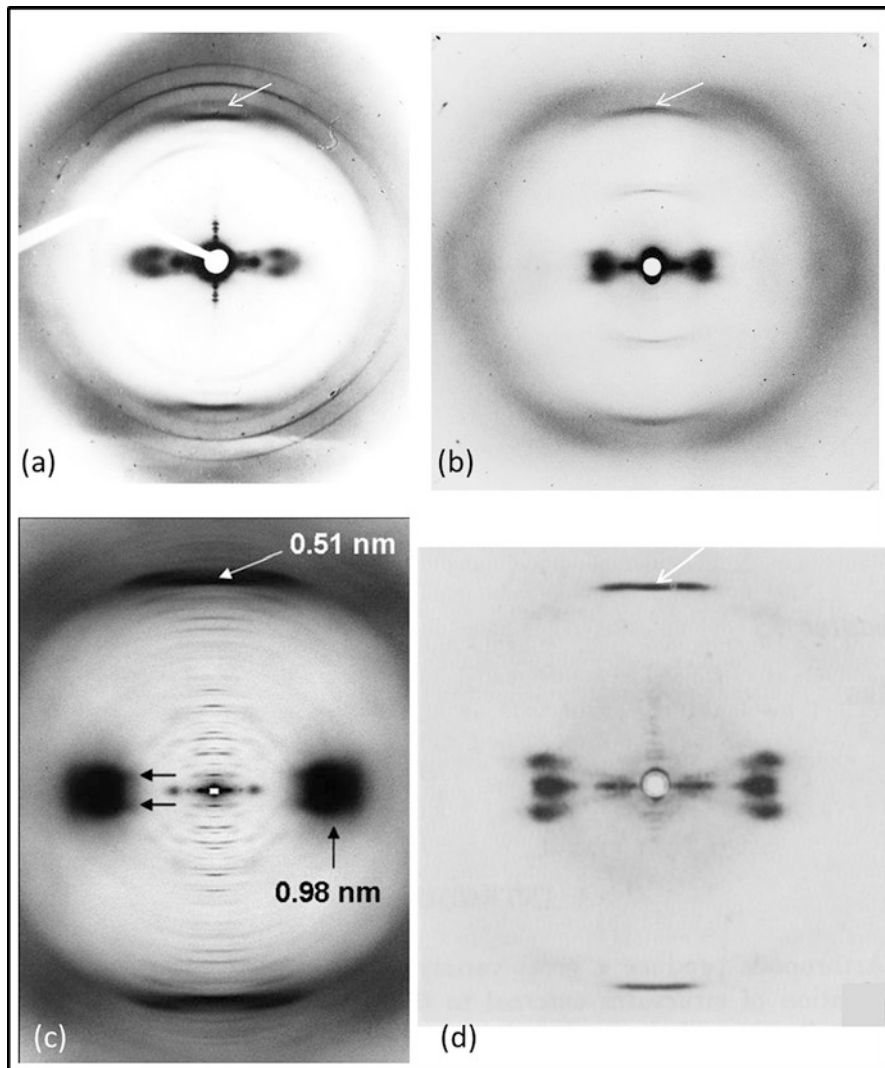
Fig. 1.3 High-angle X-ray diffraction patterns from a number of fibres with the α–helical confor-mation: (**a**) paramyosin filaments in molluscan muscle, (**b**) poly-γ-benzyl-L-glutamate, (**c**) α–kera-tin (from Parry et al. 2008; taken by T.P. MacRae) and (**d**) α–silk from the honeybee (*Apis mellifera*; Lucas and Rudall 1968)

X-ray reflection on the meridian of the diffraction pattern at a spacing of about 0.15 nm, corresponding to the axial rise per residue. Around this time there was a very strong fibre diffraction group at the Courtaulds Ltd. Research Laboratory in Maidenhead, Berkshire, UK. As part of their work, they had been studying synthetic polypeptides in which every amino acid is the same. One of major importance,
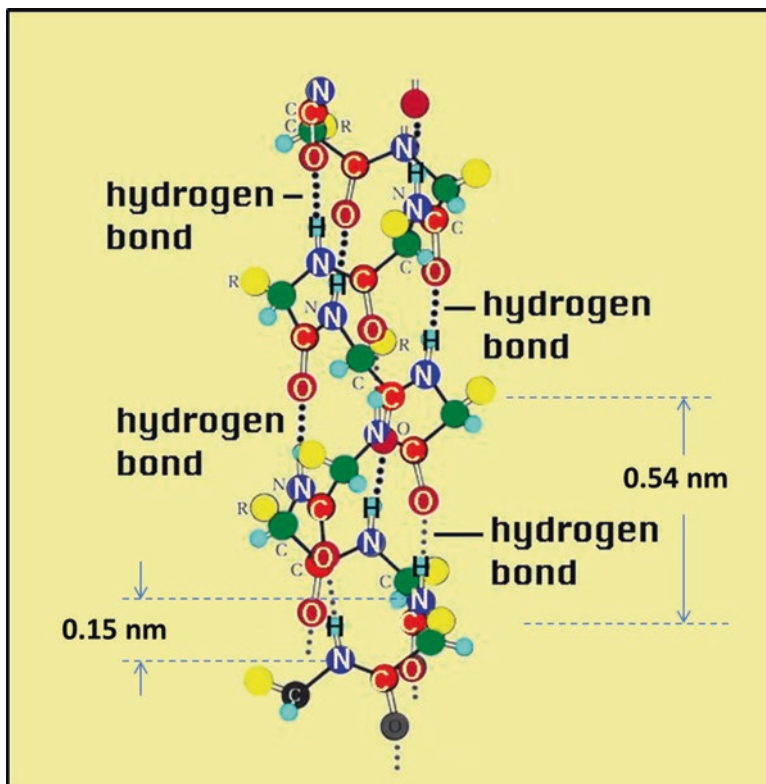
Fig. 1.4 Diagram of the α–helix structure showing the 0.54 nm helix pitch, the 0.15 nm axial translation between successive amino acids and the hydrogen-bonds that lie almost parallel to the helix axis (Adapted from http://academic.brooklyn.cuny.edu/biology/bio4fv/page/alpha_h.htm)

poly-γ-benzyl-L-glutamate (PBLG), was particularly popular and amenable for study and could be induced to form oriented films and fibres. One of our heroes from Courtaulds, Arthur Elliott (Fig. 1.2g), inventor of the toroid X-ray camera and PhD supervisor to both authors (see Squire and Vibert 1987), provided Perutz with a sample of an oriented PBLG film. Perutz tested it on a cylindrical X-ray camera with the film oscillating through a range of angles to satisfy Bragg's diffraction law (Perutz 1951). A clear 0.15 nm meridional reflection was seen, thereby providing direct experimental confirmation of the presence of the α-helix. None of the other structures proposed around that time would give such a reflection. Perutz and Huxley, in their very next paper in Nature (Huxley and Perutz 1951), showed the presence of a 0.15 nm peak in a diffraction pattern from frog sartorius muscle, confirming the presence of the α-helix in a native tissue.

1.1.3 The Coiled Coil and Heptads

Typical α-patterns from the k-m-e-f group of proteins and others are shown in Fig. 1.3. These include (a) molluscan muscle, (b) PBLG, (c) keratin and (d) α-silk from honeybee. Although the meaning of the 0.15 nm reflection is unambiguous, there are puzzles in these patterns. The pitch of the α-helix is about 0.54 nm, but the strong peaks arrowed (white) in Fig. 1.3 (a, c, d) are at a spacing of about 0.51 nm, not the 0.54 nm expected for the turn layer-line in an α-helical diffraction pattern. In addition, the region of the pattern close to the equator, where the equator is represented by the horizontal line through the middle of the patterns in Fig. 1.3, should have very little intensity. However, very clear but nonetheless unexpected peaks are observed (black arrows in (c)) just above and below the equator. These near-equatorial peaks are particularly clear in Fig. 1.3d. Details of the expected diffraction from an α–helical structure are given by Squire (1981). Why does the turn layer line have the wrong spacing and where do the near-equatorial diffraction maxima come from?

It was Francis Crick (Fig. 1.2h), later of DNA double helix fame with Jim Watson and Maurice Wilkins (and all to become Nobel Laureates), who came up with the right answer. Pauling and Corey (1953a) had suggested that particular amino acid side chain sequences could give rise to systematic distortions of the α-helices, but this time it was they who missed the main point. Crick suggested (Crick 1952, 1953b) that two α-helical molecules might twist around each other as in Fig. 1.5c so that some R-groups on one helix might fit into spaces between R-groups on the second chain. This is the so-called knob-into-hole packing (Fig. 1.5b). A feature of this packing is that the two α-helices are coiled in such a way that each has a quasi-repeat after seven amino acid residues (a heptad repeat). These residues are in defined locations, specified by the letters **a** to **g**, in what was to be termed a coiled coil structure (Fig. 1.5a, b). The particular amino acids on the line of contact between the two molecules are in the **a** and **d** positions. Crick (1953a) showed that the X-ray diffraction pattern expected from such a coiled coil would give the observed 0.15 nm meridional peak, that the 0.54 nm turn layer-line would become a 0.51 nm meridional peak, and that strong intensity would be expected just above and below the equator. The spacing of the near equatorial layer line (*P/n*) depends on the pitch length of the coiled coil (*P*) and the number of α-helical strands (*n*). A simple explanation of this pattern is shown in Fig. 1.6.

The coiled-coil explained most of the observed diffraction patterns, but a minor puzzle was the diffraction pattern from PBLG (Fig. 1.3b). We have seen that this gave Perutz his proof of the α-helical structure from its 0.15 nm meridional reflection. The PBLG pattern also had very clear near-equatorial layer lines, like the α-proteins, but on this occasion, there was also evidence that the α–helices were straight and not bent into coiled coils (Parry and Elliott 1967; Squire and Elliott 1972). It was suggested that the long sidechains in PBLG, which terminated in relatively massive benzyl groups, were producing the near-equatorial diffraction lines through specific stacking of the benzyl groups. This is an unusual case compared to
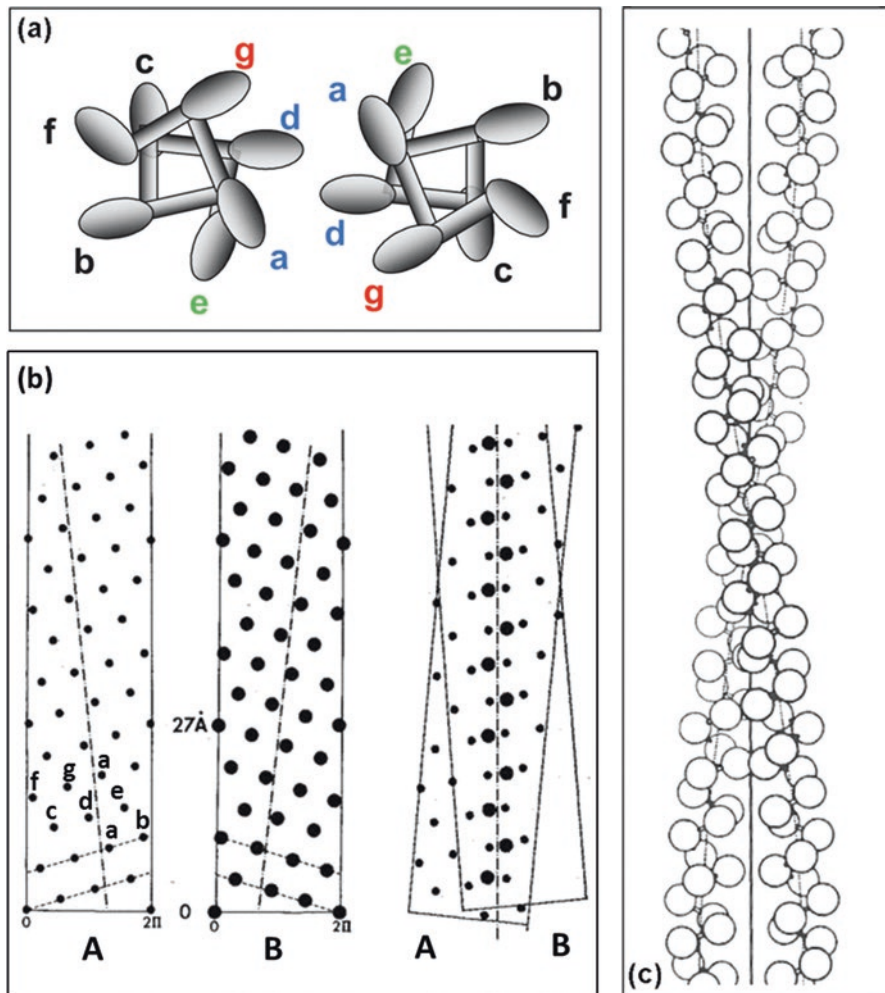
Fig. 1.5 (**a**) Schematic diagram showing the **a** to **g** heptad positions in two adjacent two-chain coiled coils. Apolar residues usually occur in positions **a** and **d**. Charged groups in positions **e** and **g** can determine the polarity of the two chains and their relative axial stagger. Most commonly, the charges in **e** and **g** are opposite and the two chains tend to be parallel. Other charge arrangements, however, can lead to an antiparallel chain arrangement. (**b**) Radial projections of the R-group positions in the two chains shown in (**a**) when viewed from right to left in (**a**). A radial projection can be thought of as being obtained by wrapping a piece of paper around the helix, marking on it the R-group positions and then unwrapping the paper. The left and right hand edges marked 0 and 2π would come together if the radial projection was refolded into a cylinder. The left hand image in (**b**) shows the left hand helix in (**a**) viewed from the *outside*. The central image shows the R-group distribution of the right hand chain in (a) viewed from the *inside*, so the hand appears opposite to that in A. The *dashed lines* in the left and central images show the line of contact of the two chains A and B when they form a coiled coil. In A this can be seen to be along the **a** and **d** positions of the heptad. The right hand image shows that the axes of the two chains need to be tilted to bring the *dashed* lines in A and B together. The two chains then mesh together with knob-into-hole packing. To maintain this interaction over a considerable axial distance the chains must twist around each other as in (**c**). (From Squire 1981; page 144). (**c**) Representation of a heptad (*left-handed*) two-chain coiled coil, where each amino acid has been represented by a *circle* (From Fraser and MacRae 1973).
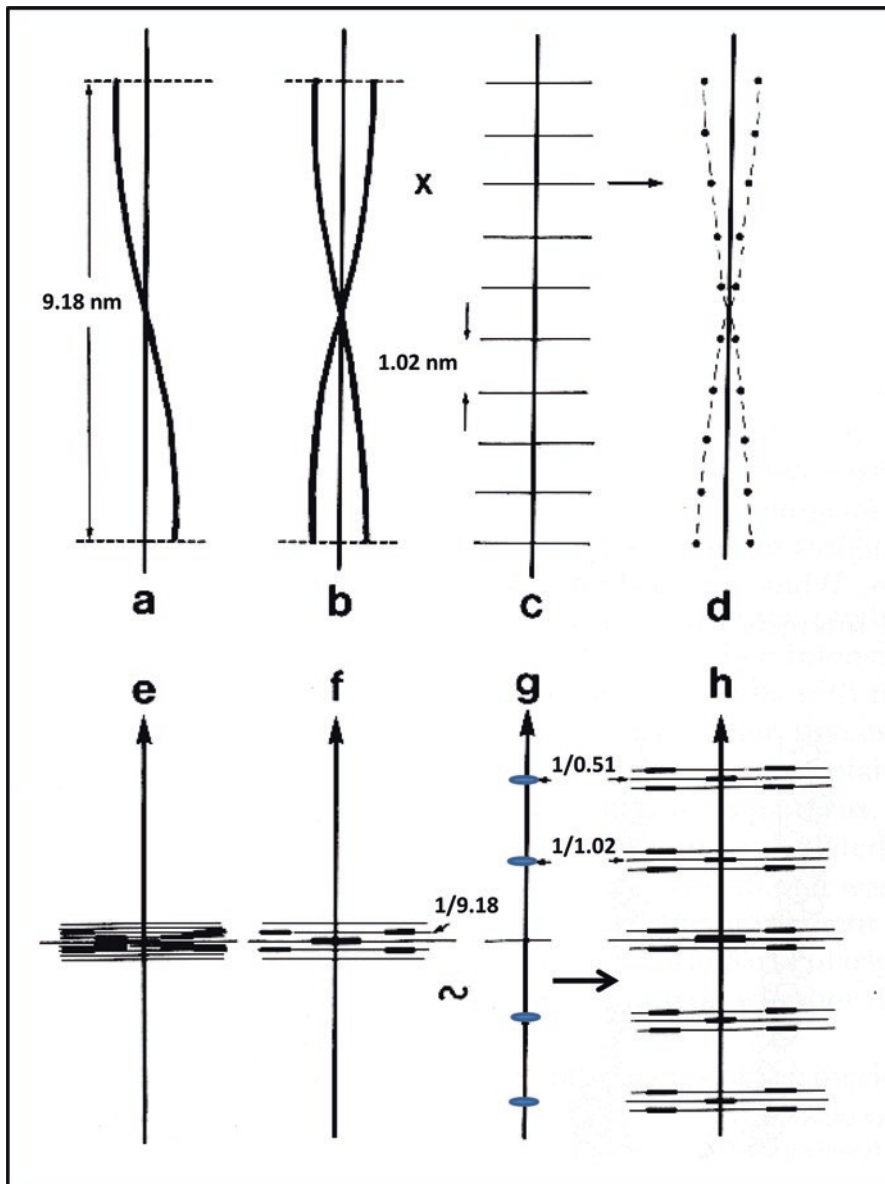
Fig. 1.6 Generation of the form of the diffraction pattern from a two-chain coiled-coil α–helical molecule. The pitch of the coiled coil is taken here to be 18.36 nm and the subunit repeat along each strand is 18.36/18 = 1.02 nm. The diffraction pattern (**e**) of a single continuous helical strand (**a**) of pitch 18.36 nm and relatively small radius (about 0.5 nm) is a very shallow helix cross with layers at positions $n/18.36$ nm^{-1} (n integer) from the equator. The introduction of a second strand coaxial with the first (**b**) halves the axial repeat to 9.18 nm and only layers for which n is even in (**e**) will be seen in the diffraction pattern (**f**). A discontinuous helix (**d**) is obtained by multiplying (**b**) by a set of planes of density that are 1.02 nm apart (**c**). The diffraction pattern (**g**) of (**c**) is a set

protein structures, but it is as well to remember that, like the α–helical backbone, the amino acid side chains will also diffract and, if systematically ordered, may contribute otherwise unexpected diffraction features (Fig. 1.7).

1.1.4 Features of the Heptad

Crick's coiled coil (Fig. 1.5c) directly explained the main features of the observed α-patterns, but what would hold the two chains together? Crick came up with the idea that residues in the **a** and **d** positions in the coiled coil might be apolar and that, in an aqueous environment, these would interact very strongly with one another. Figure 1.6 shows the structures of 20 amino acids found in proteins, where they have been grouped into families with similar characteristics. Top left are amino acids with polar sidechains. Bottom left are basic and acidic sidechains that can be ionized in an aqueous environment, depending on the pH, and which can interact with each other through their opposite charges. On the right are the apolar, non-polar or hydrophobic sidechains that will tend to cluster together if the protein is in water. Although Crick predicted that amino acids in the **a** and **d** positions of the coiled coil would tend to be apolar, there was no evidence for this assertion at that time, since sequencing of proteins had not yet proved possible. It was not until the work of Hodges et al. (1972) and Stone et al. (1975), based on the sequencing method of Sanger (1959; Fig. 1.2j), that the sequence of the first fibrous α–protein molecule was determined. This was tropomyosin, a 40 nm long coiled coil that occurs on the thin, actin-containing filaments of muscle (see Chap. 9). The sequence is shown in Fig. 1.8, with the amino acids grouped into rows of 14 residues. See Fig. 1.9 caption for the three letter amino acid codes. What is very apparent is that the residues in the **a** and **d** positions are, indeed, largely apolar, a striking confirmation of Crick's hypothesis.

1.1.5 Multi-stranded Coiled-Coils

Myosin, α-keratin and tropomyosin are all two-chain coiled coil molecules, but three-chain coiled coils were also structures described by Crick in his seminal papers (1952, 1953b). Interestingly, a variant of the latter conformation was revealed in studies of molecules such as those in the spectrin superfamily (spectrin, α-actinin,

---

Fig. 1.6 (continued) of spots along the meridian at axial positions $m/1.02$ nm^{-1} (m integer). The diffraction pattern (**h**) of (**d**) is therefore the convolution of (**f**) with (**g**). In a coiled-coil α–helix, the 1.02 nm repeat contains substructure in the form of two turns of the α–helix, each at around 0.5 nm, which effectively halves the axial repeat to 0.51 nm, together with the subunit translation of 0.15 nm. Only these regions of the diffraction pattern will be seen (i.e. at $m=2$ and $m=7$) (Taken from Squire (1981: p 147 where further details and explanation can be found))
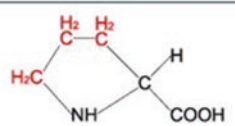
POLAR SIDE CHAINS		NON POLAR SIDE CHAINS	
SERINE	$HOH_2C - \overset{NH_2}{\underset{H}{C}} - COOH$	GLYCINE	$H - \overset{NH_2}{\underset{H}{C}} - COOH$
THREONINE	$H_3C - CHOH - \overset{NH_2}{\underset{H}{C}} - COOH$	ALANINE	$H_3C - \overset{NH_2}{\underset{H}{C}} - COOH$
TYROSINE	$HO -\!\!\bigcirc\!\!- CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$	CYSTEINE [1]	$HS - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$
ASPARAGINE	$H_2N - CO - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$	VALINE	$\underset{H_3C}{\overset{H_3C}{>}} CH - \overset{NH_2}{\underset{H}{C}} - COOH$
GLUTAMINE	$H_2N - CO - CH_2 - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$	LEUCINE	$\underset{H_3C}{\overset{H_3C}{>}} CH - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$
BASIC SIDE CHAINS		ISOLEUCINE	$H_3C - CH_2 - \overset{CH_3}{\underset{H}{CH}} - \overset{NH_2}{\underset{}{C}} - COOH$
LYSINE	$H_2N - CH_2 - CH_2 - CH_2 - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$	PROLINE	(proline ring structure)
ARGININE	$\underset{HN}{\overset{H_2N}{>}} C - \underset{H}{N} - CH_2 - CH_2 - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$		
HISTIDINE	(imidazole ring) $CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$	METHIONINE	$H_3C - S - CH_2 - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$
ACIDIC SIDE CHAINS		PHENYLALANINE	$\bigcirc - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$
ASPARTIC ACID	$HOOC - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$	TRYPTOPHAN	(indole ring) $CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$
GLUTAMIC ACID	$HOOC - CH_2 - CH_2 - \overset{NH_2}{\underset{H}{C}} - COOH$		

Fig. 1.7 The 20 amino acids that commonly occur in proteins are grouped according to their chemical properties (For details see text. Adapted from: http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/formuleAA/)

dystrophin, utrophin). It was shown that these too exhibited an α–helical coiled coil structure, but they are, nonetheless, formed from a single chain. In these cases the chain folds back on itself and then folds back again so that three parts of the same chain interact in a manner closely reminiscent of a three-chain coiled coil

g	a	b	c	d	e	f	g	a	b	c	d	e	f
1 AcMet	2 Asp	3 Ala	4 Ile	5 Lys	6 Lys	7 Lys	8 Met	9 Gln	10 Met	11 Leu	12 Lys	13 Leu	
14 Asp	15 Lys	16 Glu	17 Asn	18 Ala	19 Leu	20 Asp	21 Arg	22 Ala	23 Glu	24 Glu	25 Ala	26 Glu	27 Ala
28 Asp	29 Lys	30 Lys	31 Ala	32 Ala	33 Glu	34 Asp	35 Arg	36 Ser	37 Lys	38 Gln	39 Leu	40 Glu	41 Asp
42 Glu	43 Leu	44 Val	45 Ser	46 Leu	47 Gln	48 Lys	49 Lys	50 Leu	51 Lys	52 Gly	53 Thr	54 Glu	55 Asp
56 Glu	57 Leu	58 Asp	59 Lys	60 Tyr	61 Ser	62 Glu	63 Ala	64 Leu	65 Lys	66 Asp	67 Ala	68 Gln	69 Glu
70 Lys	71 Leu	72 Glu	73 Leu	74 Ala	75 Glu	76 Lys	77 Lys	78 Ala	79 Thr	80 Asp	81 Ala	82 Glu	83 Ala
84 Asp	85 Val	86 Ala	87 Ser	88 Leu	89 Asn	90 Arg	91 Arg	92 Ile	93 Gln	94 Leu	95 Val	96 Glu	97 Glu
98 Glu	99 Leu	100 Asp	101 Arg	102 Ala	103 Gln	104 Glu	105 Arg	106 Leu	107 Ala	108 Thr	109 Ala	110 Leu	111 Gln
112 Lys	113 Leu	114 Glu	115 Glu	116 Ala	117 Glu	118 Lys	119 Ala	120 Ala	121 Asp	122 Glu	123 Ser	124 Glu	125 Arg
126 Gly	127 Met	128 Lys	129 Val	130 Ile	131 Glu	132 Ser	133 Arg	134 Ala	135 Gln	136 Lys	137 Asp	138 Glu	139 Glu
140 Lys	141 Met	142 Glu	143 Ile	144 Gln	145 Glu	146 Ile	147 Gln	148 Leu	149 Lys	150 Glu	151 Ala	152 Lys	153 His
154 Ile	155 Ala	156 Glu	157 Asp	158 Ala	159 Asp	160 Arg	161 Lys	162 Tyr	163 Glu	164 Glu	165 Val	166 Ala	167 Arg
168 Lys	169 Leu	170 Val	171 Ile	172 Ile	173 Glu	174 Ser	175 Asp	176 Leu	177 Glu	178 Arg	179 Ala	180 Glu	181 Glu
182 Arg	183 Ala	184 Glu	185 Leu	186 Ser	187 Glu	188 Gly	189 Lys	190 Cys	191 Ala	192 Glu	193 Leu	194 Glu	195 Glu
196 Glu	197 Leu	198 Lys	199 Thr	200 Val	201 Thr	202 Asn	203 Asn	204 Leu	205 Lys	206 Ser	207 Leu	208 Glu	209 Ala
210 Gln	211 Ala	212 Glu	213 Lys	214 Tyr	215 Ser	216 Gln	217 Lys	218 Glu	219 Asp	220 Lys	221 Tyr	222 Glu	223 Glu
224 Glu	225 Ile	226 Lys	227 Val	228 Leu	229 Ser	230 Asp	231 Lys	232 Leu	233 Lys	234 Glu	235 Ala	236 Glu	237 Thr
238 Arg	239 Ala	240 Glu	241 Phe	242 Ala	243 Glu	244 Arg	245 Ser	246 Val	247 Thr	248 Lys	249 Leu	250 Glu	251 Lys
252 Ser	253 Ile	254 Asp	255 Asp	256 Leu	257 Glu	258 Asp	259 Glu	260 Leu	261 Tyr	262 Ala	263 Gln	264 Lys	265 Leu
266 Lys	267 Tyr	268 Lys	269 Ala	270 Ile	271 Ser	272 Glu	273 Glu	274 Leu	275 Asp	276 His	277 Ala	278 Leu	279 Asn
280 Asp	281 Met	282 Thr	283 Ser	284 Ile									

Fig. 1.8 The amino acid sequence of α–tropomyosin from rabbit skeletal muscle according to Stone et al. (1975). The residues have been grouped as two sets of seven and labelled with the letters **a** to **g** to show the heptad structure in which hydrophobic residues tend to occupy the **a** and **d** positions. The three letter amino acid codes are given in Fig. 1.9. (Reproduced from Squire (1981; page 186))
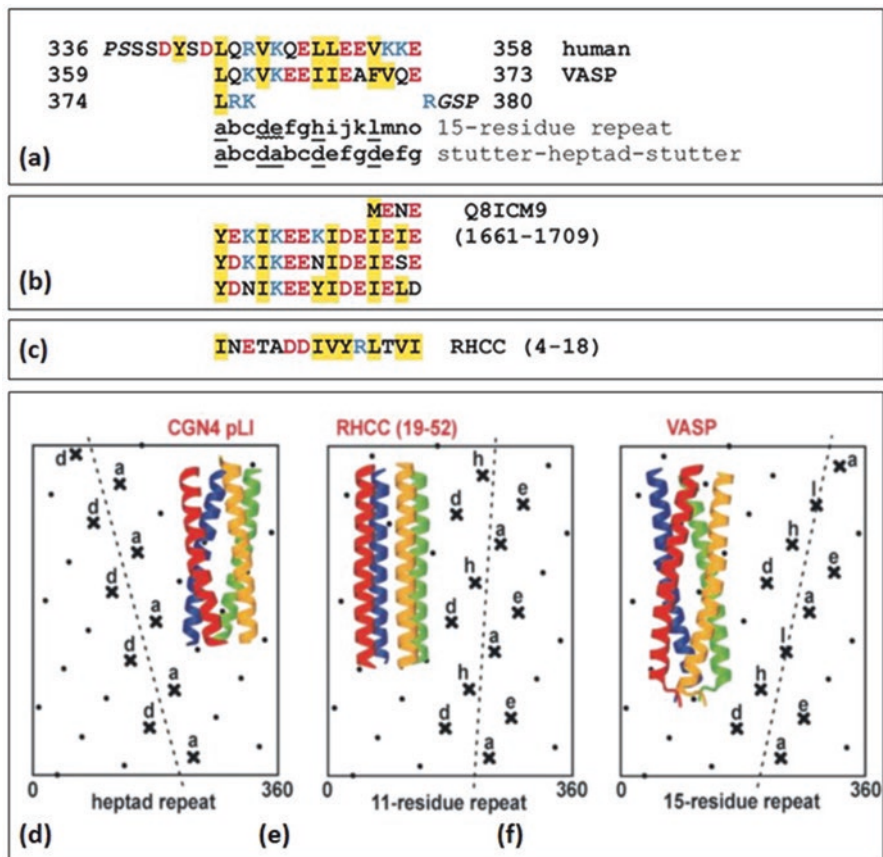
Fig. 1.9 (**a**) The amino acid sequence of the vasodilator-stimulated phosphoprotein (VASP) tetramerization domain (TD) has been formatted to highlight the 15-residue repeat pattern. Apolar residues and Y and F residues are highlighted in *yellow*, acidic residues in *red* and basic residues in *blue*. Three letter and single letter amino acid codes are **Gly**cine G, **Ala**nine A, **Val**ine V, **Leu**cine L, Isoleucine **Ile** I, **Phe**nylalanine F, **Try**ptophan (Trp) W, **Met**hionine M, **Pro**line P, **Ser**ine S, **Thr**eonine T, **Tyr**osine Y, Asparagine **Asn** N, Glutamine **Gln** Q, Cysteine **Cys** C, **Lys**ine K, **Arg**inine R, **His**tidine H, **Asp**artic acid D, **Glu**tamic acid E (See also Figs. 1.7 & 1.8). (**b**) The sequence of residues 1661–1709 of the putative myosin-like protein (TrEMBL accession no. Q81CM9) with apparent 15-residue repeats. (**c**) The sequence of residues 4–18 of a tetrabrachion fragment forming a *right-hand*, 11-residue repeat coiled-coil structure (RHCC). (**d**) Radial net diagrams (cf. Fig. 1.5b) for a general coiled coil, (**e**) 11-residue hendecad (RHCC) and (**f**) 15-residue pentadecad (VASP) repeats. The *dotted lines* show the line of contact of adjacent α-helices (Figure reproduced from Kühnel et al. (2004))

(see Chaps. 11 and 12). Long lengths of molecule can be created this way by having several repeats of this folded chain motif, known as the spectrin fold, in tandem.

Since the 1950s, coiled coils with many more strands, both parallel and antiparallel, have been discovered (Parry et al. 2008: see Chaps. 2, 3 and 4). One implication of this is that, as the number of coiled-coil strands increases, the volume of the hole in the middle of the structure so formed necessarily gets larger too, and a

solvent-occupied central channel capable of mediating solute transport is generated. Tubes of this type that have so far been identified contain five, six, ten and twelve stranded coiled coils (Chap. 4).

1.1.6 Stutters, Stammers and Coiled Coils with Specific Discontinuities

Since it became apparent that a heptad substructure in an α-helix-favouring protein would indicate a coiled coil structure, it was initially considered probable that the underlying heptad substructure would be a continuous one and not suffer phase discontinuities along its length. Indeed, when the first sequence of an α-fibrous protein (tropomyosin; Fig. 1.8) was completed by Stone et al. (1975) the heptad repeat extended from one end of the chain to the other. We now know that this feature is unusual in α-fibrous proteins and that the majority of coiled coil chains do indeed contain discontinuities in heptad phasing. Six possible types of discontinuity exist, but Brown et al. (1996) noted that the most common of these corresponded to insertions of four residues (which are structurally equivalent to deletions of three residues), and insertions of three residues (which are equivalent to deletions of four residues) in an otherwise continuous heptad substructure. These were termed stutters and stammers, respectively. The stutter results in a local unwinding of the coiled coil to give an increase in the local pitch length. The constituent α–helical strands therefore lie nearly parallel to the coiled coil axis. The structure locally is that of an 11-residue hendecad repeat comprising two heptads (14 residues) with one stutter (a three-residue deletion) (Fig. 1.9). In contrast, the stammer results in a tighter local winding of the coiled coil with a concomitant local reduction in the coiled coil pitch length. The insertion of one residue (or deletion of six residues), often termed a skip, is directly equivalent to two stutters. In effect, this can be considered as equivalent locally to a 15-residue pentadecad repeat comprising three heptads (21 residues) with two stutters (two three-residue deletions). Structurally, this results in a short piece of right-handed coiled coil (Fig. 1.9). The remaining heptad discontinuities have now all been solved crystallographically. In the cases of both the two and six residue insertions (equivalent to the five and one residue deletions, respectively), the structure of the trimeric coiled coils observed in autotransporter adhesins (for example) copes with these conformationally-significant discontinuities in heptad substructure by forming a short β-like strand that lies near-perpendicular to the axis of the coiled coil. These have been termed α/β coiled coils. More details of these conformations, as determined by Hartmann and colleagues, may be found in Chap. 3.

1.2 Beta Structures

In studies of some α-proteins, such as those from hair, it was found that if the fibres were stretched they would then give a diffraction pattern similar to those seen in some naturally-occurring silks, as in Astbury's β-pattern (Fig. 1.10). Pauling and
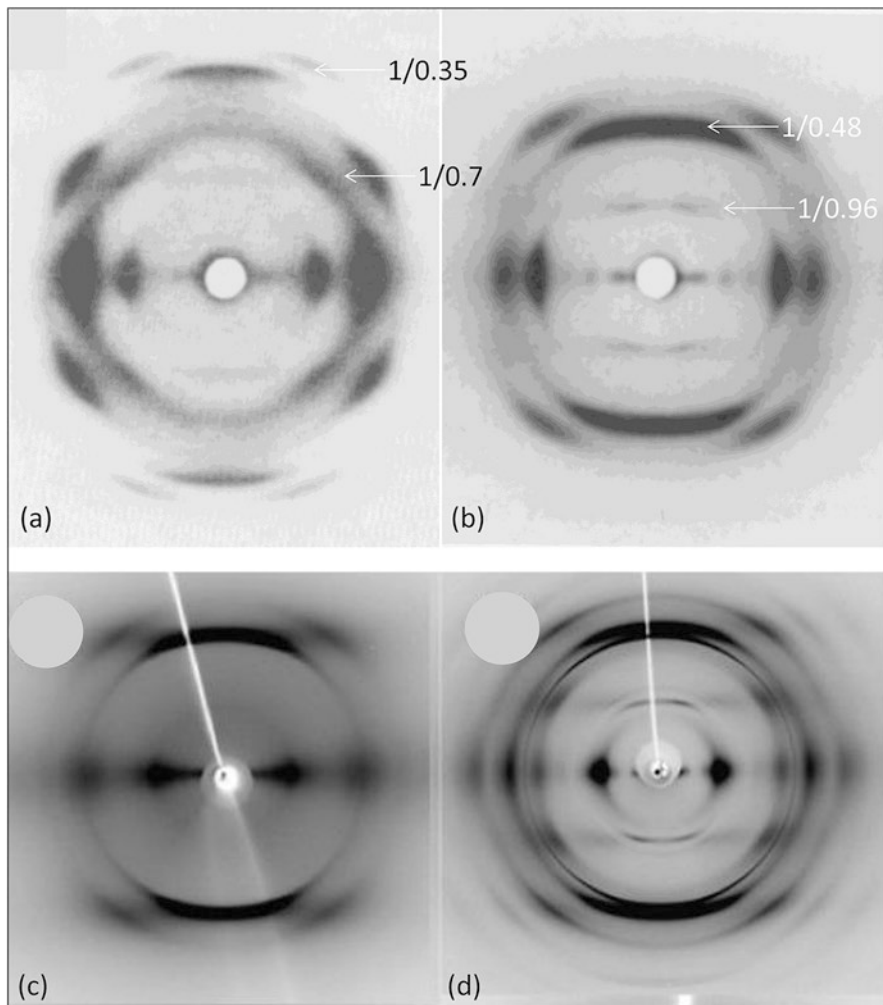
Fig. 1.10 Typical X-ray diffraction patterns from β-structures: (**a**) β-silk, (**b**) cross-β structure, (**c**, **d**) amyloid proteins. (**c**, **d**) (From Serpell (2013))

Corey (1951c, 1953b), in further ground-breaking work, discovered that extended chain polypeptide structures could be generated with systematic hydrogen-bonding between adjacent chains. The precise bonding scheme depended on whether the chains were parallel or antiparallel (Fig. 1.11a, b). In these sheets, the R-groups would project above and below the plane of the sheet created by the polypeptide backbone, but the backbone would pucker slightly to accommodate the R-groups in what is termed the β-pleated sheet. In some cases, a single polypeptide chain can fold back on itself after a few residues to give a compact antiparallel β-sheet, with the chain axes lying approximately parallel to the fibre axis. However, in less common instances, a cross-β structure may be generated instead (Fig. 1.11c), in which the chain axes lie perpendicular to the fibre axis.
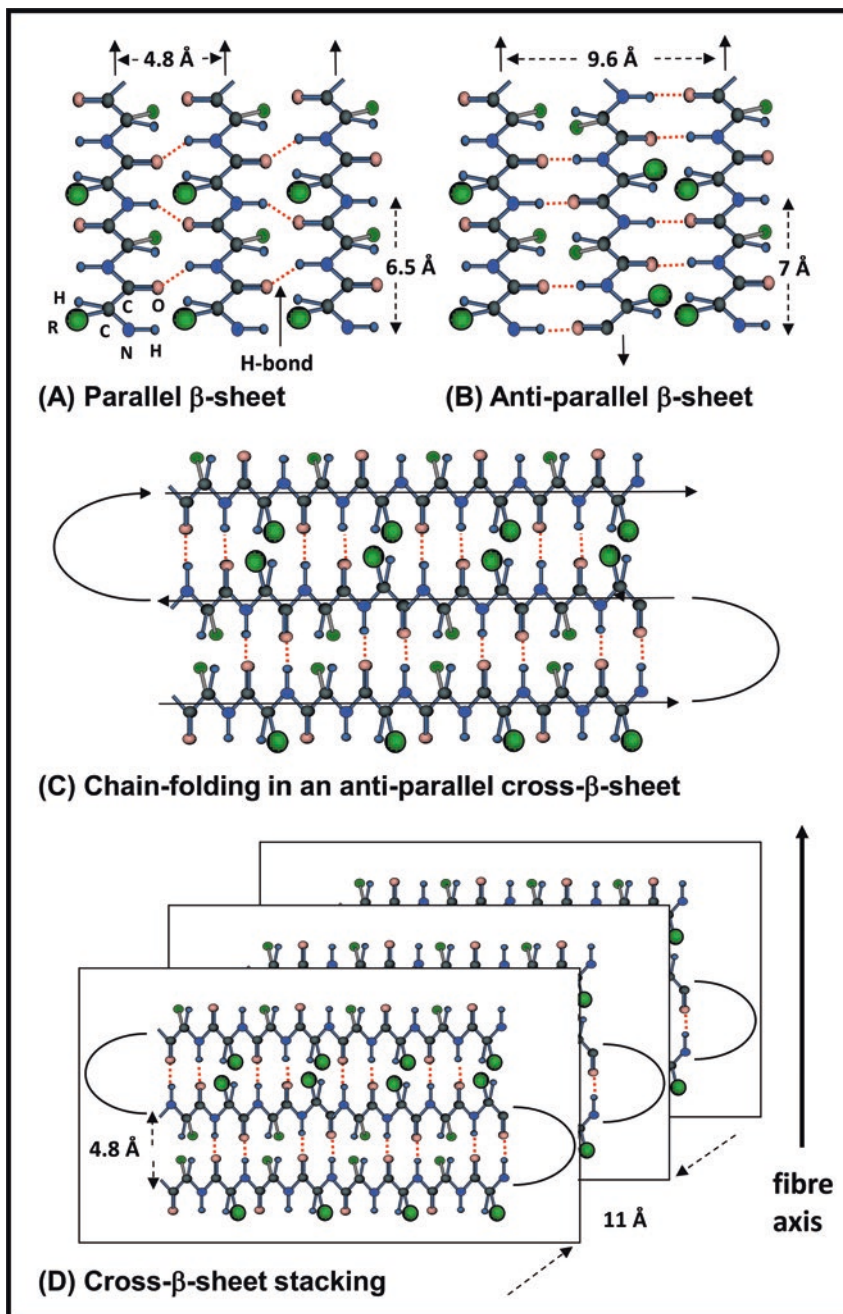
Fig. 1.11 The basic arrangements of β-strands in hydrogen-bonded β-sheets; (**a**) parallel chains, (**b**) antiparallel chains. *Green spheres* of different sizes denote sidechain groups directed either towards (*large spheres*) or away from (*small spheres*) from the reader. Hydrogen-bonds are shown by *red dotted lines*. Other colours follow the standard CPK scheme. (**c**) Chain folding back onto itself in a cross-β sheet. (**d**) Stacking of several sheets as in (**c**); the spacing of the stacks, shown as 1.1 nm, is actually very variable and depends on the nature of the R groups (From Kajava et al. (2006))

All of these extended chain structures produce sheets of molecules, and these sheets can, in turn, pack together in 3D to form extended assemblies. Figure 1.12 shows the characteristic differences between diffraction patterns (B, D) from stacks of (A) axially aligned β–chains and (C) cross-β chains. Some characteristic peaks, such as those relating to distances within individual chains, are commonly about 0.6–0.7 nm and correspond to a two amino acid repeat. Other peaks arise from distances between hydrogen-bonded chains. In the case of a parallel chain β-sheet this would be about 0.48 nm, but for an antiparallel β-sheet, the distance would be about 0.96 nm, with a strong halving due to the intervening chain of opposite polarity. In a cross-β structure, there would be a corresponding switch of the diffraction maxima from the meridian to the equator (and vice versa). The separation of adjacent sheets
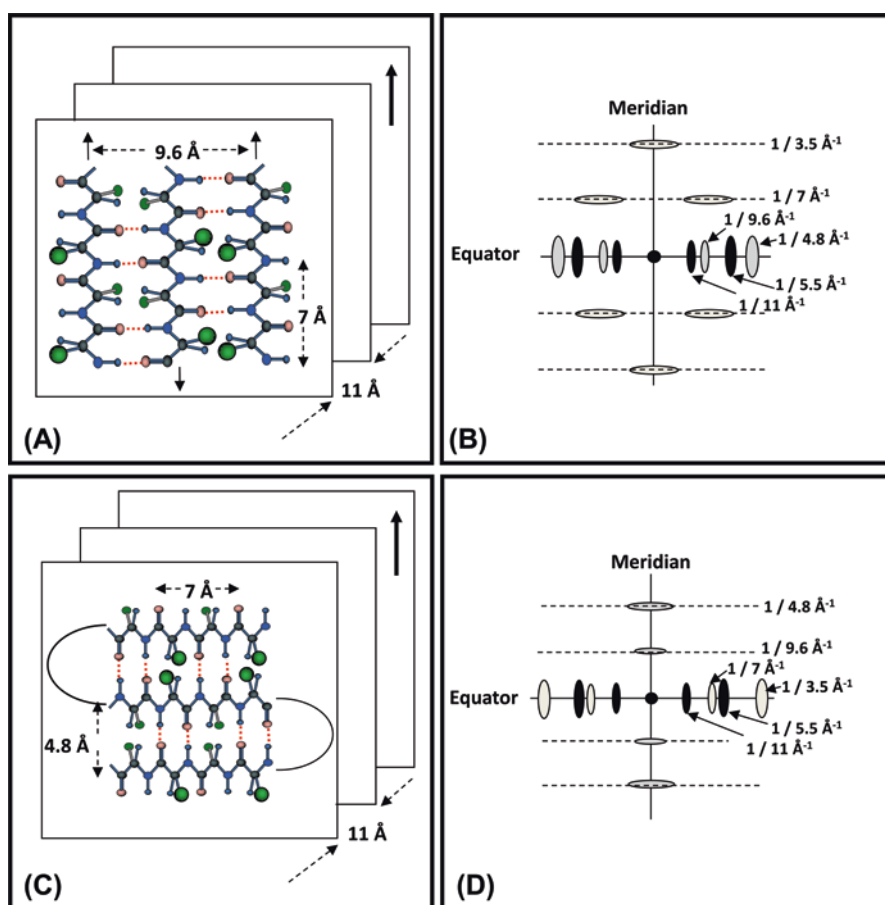
Fig. 1.12 The differences that might be observed in the fibre diffraction patterns from oriented samples of antiparallel β-structures depending on whether the chains are aligned along (**a**, **b**) or perpendicular to (**c**, **d**) the fibre axis. Colour coding on (**a**, **b**) as in Fig. 1.11. See actual X-ray patterns in Fig. 1.10 (From Kajava et al. (2006))

in protein structures in general is much more variable, however, and depends on the particular R groups involved. Figure 1.10a is from a β–silk with the chains axially aligned, and (b, c, d) are all from cross-β structures with (c) and (d) being from amyloids. Some amyloids are aberrant structures associated with medical conditions like Alzheimer's disease and so are of considerable topical interest (see Kajava et al. 2006). The structures of naturally-occurring β–keratins from the epidermal appendages of birds and reptiles are discussed in Chap. 8.

1.3   The Collagen Fold

Connective tissues, such as tendon, skin and cornea, are composed of collagen fibrils, which are, in turn, highly specific assemblies of collagen molecules. Collagens, in general, are rich in the amino acids glycine, proline and hydroxyproline (the latter two are strictly imino acids). As seen in Fig. 1.7, proline and its derivative hydroxyproline have cyclic R groups that fold back and bond with the backbone nitrogen. This rather rigid structure makes them incompatible with the formation of α–helices or β–sheets (Pauling and Corey 1951b). Early studies showed that collagen sequences usually consisted of repeating triplets of the form Gly-X-Y, with X often being proline and Y often being hydroxyproline. The structures formed by such sequences were first elucidated by studies of the synthetic polypeptides polyproline and polyglycine (for detailed discussion and references see Fraser and MacRae 1973). Some of these polypeptides fold into left-handed helices with three residues per turn. However, the full collagen structure based on the $(Gly-Pro-Hyp)_n$ sequence has three such chains coiling around each other in a right-handed manner to give a compound structure with 10 residues in three turns (Fig. 1.14). The axial repeat is about 2.9 nm and the axial rise per subunit is 0.29 nm. The latter gives a strong meridional peak (Fig. 1.13b), and the helix pitch length of about 0.98 nm results in a very strong turn layer line. The collagen structure is stabilized by hydrogen-bonding between the glycines of the three interacting chains, which lie on the inside of the 10/3 helical structure.

Early understanding of collagen structure indicated that the Gly-X-Y repeating structure must be a fundamental feature of all collagens, and indeed some fibril-forming collagens have in excess of 1000 repeats in tandem (see Wess 2005 and Chap. 14). There are, however, other types of collagen (of the 28 types in humans) that have much more complicated structures. These contain relatively few classical triplets, many have breaks between triplets, and some form networks or other assemblies (see Knupp and Squire 2005 for a discussion of network-forming collagens).
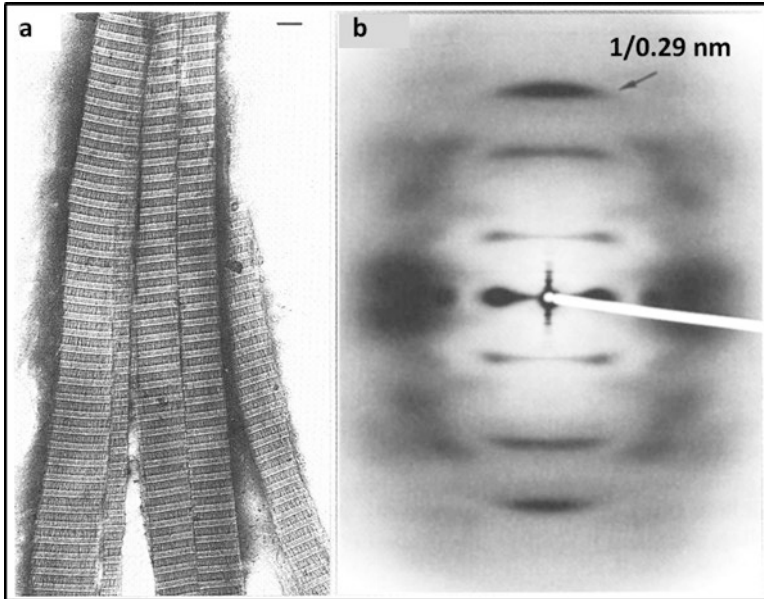
Fig. 1.13 (**a**) Transmission electron micrograph of negatively-stained cryo-sections of rat tail tendon showing the characteristic banding pattern of collagen. Scale bar is 0.1 μm. (**b**) High-angle X-ray diffraction pattern of kangaroo tail tendon that had been air-dried and then rehumidified to 75 % relative humidity (From Fraser et al. (1979))

1.4 Assembly of Building Blocks

1.4.1 Globular Proteins

We have seen that relatively simple amino acids can produce a wide variety of structures, as first determined by studies of fibrous proteins in the early 1950s. At around the same time, protein crystallography was in its early stages of development, but researchers such as John Kendrew, Max Perutz and Dorothy Crowfoot Hodgkins were actively studying crystals of interesting proteins like myoglobin, haemoglobin and insulin. They noticed some features of the crystal patterns that were similar to those seen in some fibrous protein patterns (Perutz 1949). This was the first indication that globular proteins might also contain some regularly folded regions of polypeptide chain. When these structures were eventually solved, it was found, of course, that the main features of the structures were indeed short segments of α-helix, β-sheet and some regular bends all folded together to give a compact structure. Often the apolar residues occurred predominantly on the inside of the protein, thereby helping to maintain the folded globular structure. Even the 3_1 helix of Bragg et al. (1950) was sometimes found.

Cohen and Parry (1986, 1990) (see Fig 1.2l) proposed that the coiled-coil structure, as characterised by the presence of heptad repeats, would not therefore be

Fig. 1.14 Schematic diagram showing the levels of structural organization in collagen fibrils. (**a**) the Gly-Pro-Hyp or Gly-X-Y amino acid repeat along the collagen polypeptide chain. (**b**) The fold of one chain as in (**a**) into a 3_1 helix. (**c**) The supercoiling of a 3_1 helix due to its interaction with two other similar chains (**d**) to make a 10_3 molecular helix. As well as the typical Gly-Pro-Hyp repeat along the collagen chains there are specifically placed charged and other amino acids (**e**) that make adjacent triple-helical collagen molecules interact with a 67 nm axial stagger (**f**). The Type I collagen molecules are much longer (each is about 300 nm long) than the axial stagger between molecules leading to a well-developed gap-overlap structure of the collagen fibril, as seen in Fig. 1.13a
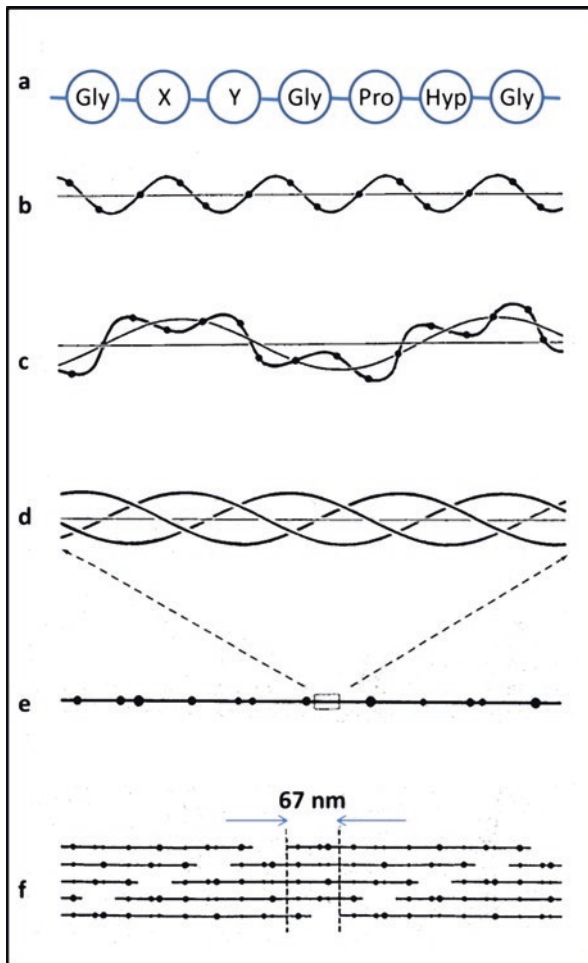
confined to the fibrous proteins, but would be a feature of globular structures too. The wealth of structures solved by protein crystallography and NMR methods has now revealed that these predictions were indeed correct, and that many globular proteins do indeed contain coiled-coil regions. Globular proteins may also contain β-sheets and yet others have short regions of a collagen-like fold. Of course, it is the amino acid sequences that define these structures, so study of the sequence of a protein to solve an unknown structure can often be helped by searching for heptads or triplet repeats.

This book is mainly about fibrous proteins, so the next few sections introduce ideas about how coiled coils, β-sheets or triple-helical collagen molecules pack together to give the extended structures that are observed in nature.

1.4.2   *Packing of Coiled Coils*

Some of the members of the k-m-e-f group of fibrous proteins have their molecules packed together to produce much larger filaments. Myosin molecules, which assemble to form the myosin filaments in muscle, and intermediate filament proteins, which form neurofilaments, trichocyte and epidermal keratin and so on, are all two-chain α–helical coiled-coil molecules. So what is their preferred mode of packing? Early ideas came from X-ray diffraction studies of muscle, particularly from molluscan muscles, where the myosin–containing filaments are very large and are mostly composed of a myosin-like protein called paramyosin, which is yet another two-chain coiled-coil. The equatorial and near-equatorial regions from the X-ray diffraction patterns of the α–fibrous proteins (Fig. 1.3) represent the regions where most of the information about molecular packing is contained. But, unfortunately, it is rarely rich in information, because native α–proteins commonly display lateral disorder. The axial spacing of the near-equatorial layer line is related to the pitch length of the coiled-coil. There is also a strong equatorial peak at a spacing of about 1 nm, and often there is additional intensity closer to the centre of the pattern that indicates that the coiled-coils pack at a centre-to-centre distance of about 2 nm. In most cases there is little further information. However, with suitable treatment, the paramyosin/myosin assemblies in molluscan muscles (see Chap. 11) can be induced to form more regular structures as in Fig. 1.15a from Elliott et al. (1968) and in Elliott and Lowy (1970). Here the equator and near-equatorial regions can be seen to be sampled at positions marked by the letters a, b and c. These are in a position that is consistent with a body-centred tetragonal structure (Fig. 1.15c), where the unit cell side is about 2.7 nm and the central molecule is shifted up the axis by half of the c-axis (axial) repeat of the unit cell. This repeat, at this low resolution, corresponds to half of the coiled coil pitch length (P), so the molecules are axially shifted by one quarter of the pitch. A body-centred unit cell gives rise to a selection rule in the observed reflections, such that $h + k + l = 2 N$ (N integer) where h, k and l are the Miller indices of the particular peak. This means that on the equator ($l = 0$), only reflections where $h + k$ is $2 N$ (i.e. even) are seen and on the 1st layer line (l=1) only reflections with $h + k = 2 N + 1$ (odd) are seen (Fig. 1.15b), as observed in Elliott et al. (1968) and Elliott and Lowy (1970).

The lateral separation of the molecules in the centred unit cell in Fig. 1.15c is about 1.9 nm. The molecular width of the coiled coil viewed from a given direction grows and shrinks as the path of the helix of constant radius rotates around the fibre axis. The P/4 axial stagger in the centred unit cell, illustrated in the inset in Fig. 1.15a, therefore makes the wider parts of one coiled-coil fit against the narrower parts of the adjacent molecules, a favourable, relatively close-packed arrangement. It was Rudall (1965), however, who had been the first to suggest that packing optimisation between two-stranded coiled coils would occur when the molecules had a relative axial stagger of P/4. Interestingly, this same axial stagger has now been observed in other two-stranded α-fibrous proteins, including those in the egg case of the praying mantis (Bullough and Tulloch 1990) and in trichocyte keratin (Fraser and Parry 2014a). A P/8 axial stagger for four-stranded coiled coils, that likewise optimizes molecular packing, is seen in the multimerisation domain of the
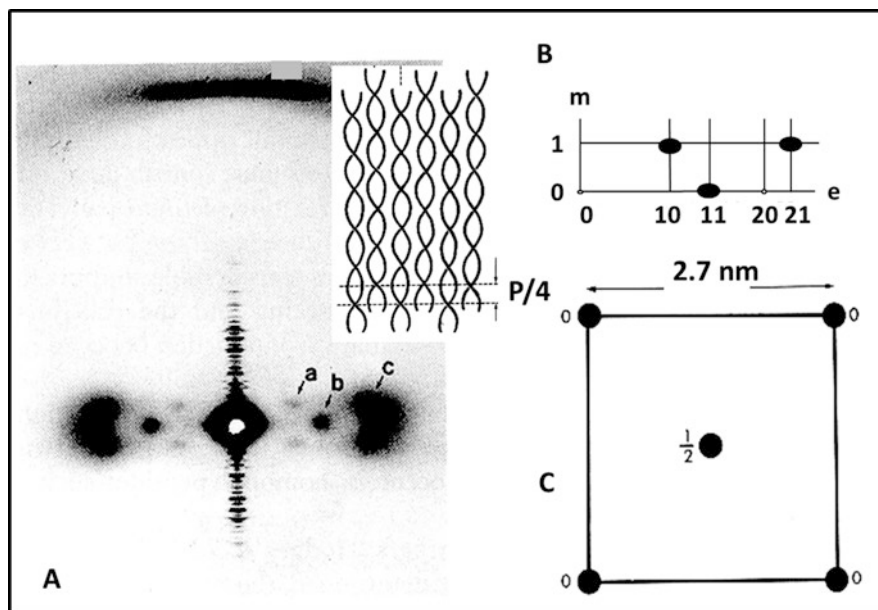
Fig. 1.15 The packing of paramyosin molecules in the thick filaments of a molluscan muscle. (**a**) X-ray diffraction pattern from *Crassostrea angulata* opaque adductor muscle. The strong 0.51 nm meridional peak characteristic of a coiled coil is at the *top*. Inset: the quarter-pitch (*P*/4) shifted stacking arrangement of paramyosin coiled coils showing the close packing thereby obtained. (**b**) The expected sampling of the equator and the near-equatorial layer lines from a body-centred tetragonal structure as in (**c**). For details and the significance of *a*, *b* and *c* in (**a**) see text (Figure based on Elliott et al. (1968) and Elliott and Lowy (1970))

four-stranded coiled coil structure in measles virus phosphoprotein (Blocquel et al. 2014) and in the α-silks from the *Hymenoptera aculeata* (Fraser and Parry 2015). It would appear, therefore, that a relative axial stagger of *P*/2*n*, between *n*-stranded coiled coils of pitch length *P*, in order to optimize molecular packing, is a common feature in both filaments and other molecular assemblies.

Chapters 5 and 6 outline the structures and properties of native intermediate filaments, and Chap. 7 describes the pathologies of intermediate filament proteins studied through the use of transgenic animals. The ways in which myosin molecules are packed to give myosin filaments are discussed in detail in Chap. 11. Tropomyosin, an important α-protein involved in the regulatory mechanism of contraction is also described (Chaps. 9 and 11).

1.4.3 Assembly of β-Sheets

Fibrous proteins containing β-sheets most commonly gain stability by assembling the sheets in a layered manner into a unit of structural/functional significance. Early sequence data from silks showed that several of them contained two-residue repeats.
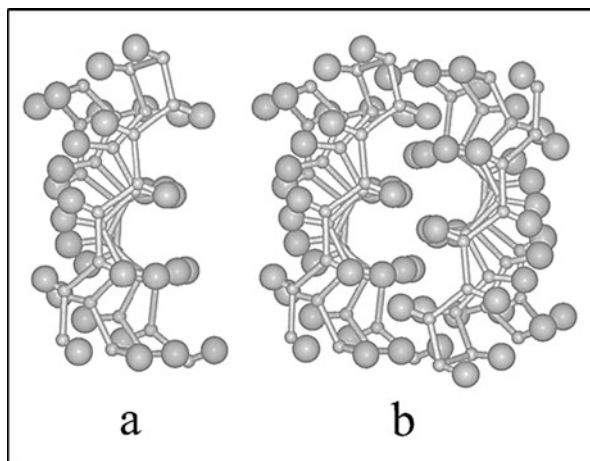
Fig. 1.16 (**a**) A β-sheet distorted to give a *right-handed* twist. The view is in the direction of the hydrogen-bonds between the chains, (**b**) a pair of these twisted β-sheets related by a perpendicular dyad axis of rotation showing the interleaving of the sidechains in the interior along the filament axis. For the β-keratins of the epidermal appendages of birds and reptiles the faces that interact are strongly hydropathic. Assembly is also facilitated by shape complementarity between the two sheets (Figure from Fraser and Parry (2008))

For example, a large fragment of the sequence of *Bombyx mori* silk is characterized by multiple, contiguous Gly-X repeats. In a β-sheet, therefore, glycine would appear on one face of the sheet and X would appear on the other face. The β-sheets would assemble with their glycine faces together, thereby stabilizing the structure in an aqueous environment. Further sheets would then assemble through the X faces, and so on. It should be noted here that β-sheets are rarely planar, but are generally twisted in a right-handed manner. Thus, shape complementarity also plays an important part in the successful assembly of two β-sheets. Of course, many fibrous proteins do not have such a simple amino acid composition as the silks. In these proteins there is not a simple set of residues on each face. A classic example is the structure of the β-keratins from the epidermal appendages of birds and reptiles. These have been shown to form 3.4 nm diameter fibrils. Chains from all species studied to date contain a highly conserved 34-residue segment that forms a four- to five-stranded, antiparallel β-sheet connected by tight turns. One face of this sheet, as modelled by Fraser and Parry (2014b) on the basis of the known sequences, is strongly hydropathic (apolar), whereas the other face contains most of the charged and hydrogen-bonding residues. It would seem self-evident that the hydropathic faces would interact with one another, and that these would be related by either a parallel or perpendicular dyad axis. Furthermore, the hydropathic faces would be located internally. The second face, of course, with its more charged residues and greater hydrogen-bonding capability would be directed outwards to the aqueous environment (Fig. 1.16). Again, the shape complementarity of the β-sheets would be expected to play an important part in their assembly. The structure of these two-sheet filamentous β-keratins is given in detail in Chap. 8.

1.4.4   Collagen Fibrils

We have seen that collagen chains fold to give the triple-helical structure illustrated in Fig. 1.14. This is the so-called Rich – Crick II conformation (Fig. 1.2i; Rich and Crick 1961), and was subsequently confirmed in an elegant study undertaken by Fraser et al. (1979), in which high-resolution X-ray diffraction data were used in conjunction with a linked-atom least-squares analysis to refine the molecular conformation. But how do these molecules pack to produce collagen fibrils, which are the main functional units in many connective tissues? Typical fibrils of Type 1 collagen are shown in Fig. 1.13a, where a very clear axial banding pattern can be seen using electron microscopy or deduced from the meridional X-ray diffraction patterns. The axial repeat, often referred to as the D-period, is about 67 nm, depending on, for example, the state of hydration and on the particular source of collagen. The long collagen molecules (each about 300 nm in length), have linear periodicities of charged, apolar and other amino acids along their lengths superimposed on the predominant Gly-Pro-Hyp substructure (Hulmes et al. 1973). Hodge and Petruska (1963) suggested that the 67 nm period observations were best explained by arranging the molecules with an axial stagger of D (or a multiple). Later, it was shown by Hulmes et al. (1973) that this arrangement is specified by both ionic and apolar interactions between residues in different molecules, where clear maxima in the numbers of interactions peak at axial distances of D and multiples thereof (Fig. 1.14f). Covalent crosslinks between molecules also contribute significantly to the structural integrity of the fibrils. It is also pertinent to note that a multiple of D (67 nm) does not correspond to the molecular length (300 nm), and hence the fibril will necessarily display a gap-overlap substructure.

How do these collagen molecules pack in three-dimensions? The high-angle diffraction patterns in Figs. 1.13b and 1.16 show fanning layer lines and a fan-shaped equatorial intensity (double arrows in Fig. 1.16). These features are indicative of the collagen molecules being slightly tilted with respect to the fibre axis. However, with very careful preparation, the lower angle collagen X-ray pattern (Fig. 1.16) can be seen to feature strong, finely-spaced meridional peaks, orders of the 67 nm repeat, together with fine sampling along (near vertical) row-lines on each side of the meridian. In addition, there is also sampling of the strong turn layer line with a spacing of 0.98 nm. Bruce Fraser (Fig. 1.2h), another of the heroes of the fibrous protein story, solved the structure in some considerable detail (Fraser et al. 1983, 1987) in what is considered as a veritable tour-de-force by those in the field. In essence, the conformation is that of highly-ordered straight segments of tilted collagen molecules in the overlap region linked to one another by relatively disordered molecular segments that constitute the gap regions. Thus, the concept of a D-periodic crimp is an inescapable conclusion from the X-ray data (see Parry 1988 for a summary of the molecular and fibrillar structure of connective tissue). The fine sampling discussed above has also been modelled by Orgel et al. (2001) to reveal further details of the molecular packing as described in Chap. 14 (Fig. 1.17).

Fig. 1.17 The central part of the X-ray diffraction pattern from a well-ordered sample of collagen showing the fan-shaped equatorial intensity (*double arrow*) and the fine sampling of the meridian, various row-lines and the turn layer line at 0.98 nm. These features have been analysed by Orgel et al. (2001) to give a detailed three-dimensional model of collagen packing (Adapted from Orgel et al. (2014))
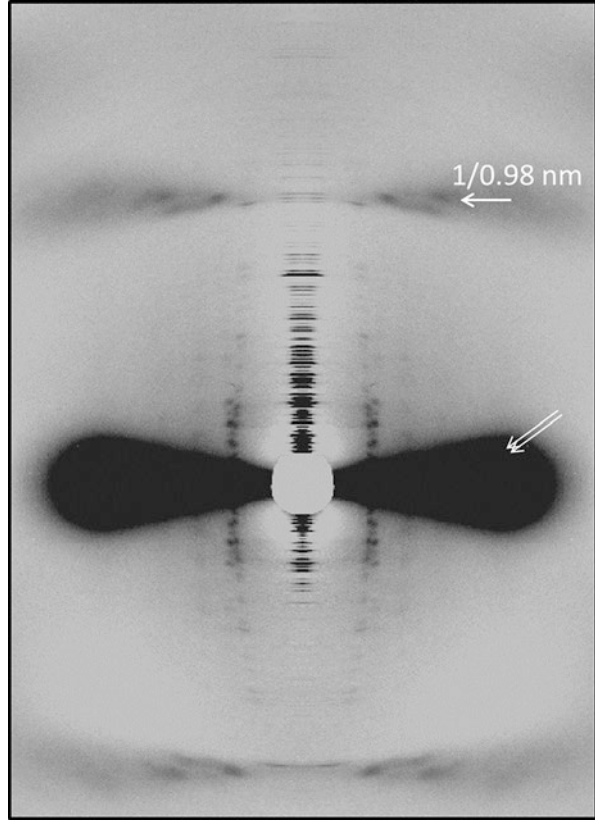
1.5 New Approaches to Solving Fibrous Protein Structures

Analysis of X-ray diffraction patterns from any material is both important and informative, but X-ray patterns from fibrous proteins often show the presence of considerable disorder (frequently lateral rather than axial), which precludes detailed interpretation. It is, of course, of considerable help to the researcher if they can also actually visualize the structures in question, and it is here that electron microscopy (TEM, transmission electron microscopy; SEM, scanning electron microscopy; STEM, scanning-transmission electron microscopy) has come into its own. For many years, in the study of fibrous proteins, X-ray diffraction and electron microscopy were considered as complementary techniques, each with its own unique advantages. X-ray diffraction patterns were recorded, often to high resolution, and sometimes from native tissue in its normal physiological states, but the problem of interpretation was formidable. On the other hand, electron microscopy of such materials has also been of great interest in that direct images can be obtained from these tissues but, until recently, the resolution has been limited to a few nm. In addition, in many cases, there has been clear evidence of radiation damage caused by the

electron beam and, furthermore, the problem of artefacts being introduced by the preparation techniques was not insignificant.

At the time of writing this Chapter these limitations on electron microscopy are gradually beginning to disappear. For some time now it has been known that the use of frozen specimens greatly improves the available resolution and better retains the native structure of the tissue being studied. In addition, electron microscope technology for studying cryo- samples and keeping them stable has greatly improved (Kühlbrandt 2014), with much emphasis on automation and the need for new users to get up to speed as quickly as possible. In addition to new improved microscopes, new detectors are being used which detect electrons directly. For many decades, electron microscope images were recorded on film. Later, detectors were used which converted the electron signal to photons which were, in turn, detected through a lens or fibre optic coupling by CCD or CMOS sensors. Direct detection of the electrons would be better, of course, but CCD and CMOS sensors were prone to radiation damage, and this had a detrimental effect on both detector performance and lifetime. To quote Clough et al. (2013) "Direct detection is an attractive alternative where signal in the CMOS or CCD sensor is generated directly by the primary electron beam." And then "The use of enclosed-gate layout transistors (ELT) and thin gate oxide improves the radiation hardness of devices, which are now attractive as a technology for electron imaging. There are a number of fundamental disadvantages to indirect detection. Scattering of the primary electrons in the scintillator leads to generation of photons in a volume larger than a single pixel. This implies that each electron generates signal in a cluster of pixels on the sensor. The effect of an electron being detected as signal in multiple pixels can be described quantitatively using the modulation transfer function (MTF). Direct detectors offer the possibility of significantly improving this MTF. Of particular interest are thin detectors where much of the inactive supporting substrate has been removed, leaving a thin active layer. Signal is generated in the active layer and then the electrons exit the active layer before significant lateral scattering has occurred. The removal of the supporting substrate prevents electrons scattering from the substrate back into the active layer, which considerably improves MTF (McMullan et al. 2009, 2014)".

Improvements in the microscope and detectors are not the only factors involved in the efficacy of modern electron microscopy. Ways of processing the images have also improved greatly in the last decade and a half. For several years it has been known that electron micrographs of images of the same kinds of particle viewed in different directions in the microscope could be sorted and averaged to provide details of the object in 3D using a type of tomographic reconstruction process. This is a method called single particle analysis (Subramaniam et al. 2016). In these cases the particles have often been globular proteins or protein assemblies held in a thin layer of ice, and many thousands of images of the same particles have been put together using programs such as IMAGIC (van Heel et al. 2000), Spider (Frank et al. 1996) and EMAN (Ludtke et al. 1999) to produce three-dimensional density maps of the particle at spectacular resolutions, sometimes around 0.3 to 0.4 nm. This method is starting to compete favourably with protein crystallography that, together with high field NMR, used to be thought of as the only techniques available

to solve structures to high resolution. Single particle analysis is not restricted to globular particles. Many of the fibrous proteins of interest are long filaments with a regular axial repeat. Images of these filaments can be chopped into small segments containing one or more repeats and these segments then become the single particles to be averaged (Egelman 2000; AL-Khayat et al. 2004; Paul et al. 2004, 2010). Some examples of the application of single-particle analysis methods are given in Chap. 11.

1.6   The Future: Protein Engineering

Chapters 2, 3 and 4 describe what is known about coiled coils and outline relatively new departures in the fibrous protein story. From a host of studies of the factors affecting coiled-coil structures, rules have been discovered which relate the protein sequence to the particular kind of coiled coil that is produced. This means that, in principle, if it is desired to generate a coiled coil of a certain structure, a peptide can be synthesised with the appropriate sequence. This is a form of protein engineering, and Chap. 2 outlines what has been achieved to date. Chapters 15, 16, 17 and 18 add significantly to this story in that they describe what can now be done with recombinant proteins, silk proteins, collagen proteins and further coiled coil peptides to generate new biomaterials. These are of use in, for example, medical applications, such as drug delivery and tissue replacement. A whole new field of discovery and exploitation is opening up. The story of the adventures of the humble amino acid is far from over.

References

AL-Khayat HA, EP Morris, JM Squire (2004) Single particle analysis: a new approach to solving the 3D structure of myosin filaments. J Muscle Res Cell Motil 25:635–644
Astbury WT (1949a) Structures of alpha-keratin and synthetic polypeptides. Nature 164:439
Astbury WT (1949b) The structure of biological tissues as revealed by X-ray diffraction analysis and electron microscopy. Br J Radiol 22:355–365
Astbury WT (1951) Adventures in molecular biology. Harvey Lect. 1950–1951 Series 46:3–44
Astbury WT, Bell FO (1941) Nature of the intramolecular fold in α-keratin and α-myosin. Nature 147:696–699
Astbury WT, Dickinson S (1940) X-ray studies of the molecular structure of myosin. Proc R Soc Lond B 129:307–332
Astbury WT, Woods HJ (1930) The X-ray interpretation of the structure and elastic properties of hair keratin. Nature 126:913–914

Astbury WT, Woods HJ (1933) X-ray studies on the structure of hair, wool and related fibres II: the molecular structure and elastic properties of hair keratin. Phil Trans R Soc Lond A 232:333–394

Atkins EDT (1967) A four-strand coiled-coil model for some insect fibrous proteins. J Mol Biol 24:139–141

Blocquel D, Habchi J, Durand E, Sevajol M, Ferron F, Erales J, Papageorgiou N, Longhi S (2014) Coiled-coil deformations in crystal structures: the *measles virus* phosphoprotein multimerization domain as an illustrative example. Acta Crystallogr D70:1589–1603

Bragg L, Kendrew JC, Perutz MF (1950) Polypeptide chain configurations in crystalline proteins. Proc R Soc A 203:321–357

Brown JH, Cohen C, Parry DAD (1996) Heptad breaks in α-helical coiled coils: stutters and stammers. Proteins Struct Funct Genet 26:134–145

Bullough PA, Tulloch PA (1990) High-resolution spot-scan electron microscopy of microcrystals of an α-helical coiled-coil protein. J Mol Biol 215:161–173

Clough RN, Moldovan G, Kirkland AI (2013) Direct detectors for electron microscopy. J Phys Conf Ser 522:012046

Cochran W, Crick FHC, Vand V (1952) The structure of synthetic polypeptides. 1. The transform of atoms on a helix. Acta Crystallogr 5:581–586

Cohen C, Parry DAD (1986) α-Helical coiled-coils – a widespread motif in proteins. Trends Biochem Sci 11:245–248

Cohen C, Parry DAD (1990) α-Helical coiled-coils and bundles: how to design an α-helical protein. Proteins Struct Funct Genet 7:1–15

Corey R, Donohue J (1950) Interatomic distances and bond angles in the polypeptide chain of proteins. J Am Chem Soc 72:2899–2900

Crane HR (1950) Principles and problems of biological growth. Sci Mon 70:376–389

Crick FHC (1952) Is alpha-keratin a coiled coil? Nature 170:882–883

Crick FHC (1953a) The Fourier transform of a coiled-coil. Acta Crystallogr 6:685–689

Crick FHC (1953b) The packing of α–helices: simple coiled coils. Acta Crystallogr 6:689–697

Egelman E (2000) A robust algorithm for the reconstruction of helical filaments using single particle methods. Ultramicroscopy 85: 225–234

Elliott A, Ambrose EJ (1950) Evidence for chain folding in polypeptides and proteins. Faraday Soc Disc 9:248–251

Elliott A, Lowy J (1970) A model for the coarse structure of paramyosin filaments. J Mol Biol 53: 181–203

Elliott A, Lowy J, Parry DAD, Vibert PJ (1968) Puzzle of the coiled coils in the α–protein paramyosin. Nature 218:656–659

Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and DWEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116:190–199

Fraser RDB, MacRae TP (1973) Conformation in fibrous proteins. Academic Press, New York/London

Fraser RDB, Parry DAD (2008) Molecular packing in the feather keratin filament. J Struct Biol 162:1–13

Fraser RDB, Parry DAD (2014a) Keratin intermediate filaments: differences in the sequences of the type I and type II chains explain the origin of the stability of an enzyme-resistant four-chain fragment. J Struct Biol 185:317–326

Fraser RDB, Parry DAD (2014b) Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties. J Struct Biol 188:213–224

Fraser RDB, Parry DAD (2015) The molecular structure of the silk fibers from *Hymenoptera aculeata* (bees, wasps, ants). J Struct Biol 192:528–538

Fraser RDB, MacRae TP, Suzuki E (1979) Chain conformation in the collagen molecule. J Mol Biol 129:463–481

Fraser RDB, MacRae TP, Miller A, Suzuki E (1983) Molecular conformation and packing in collagen fibrils. J Mol Biol 167:497–521

Fraser RDB, MacRae TP, Miller A (1987) Molecular packing in Type I collagen fibrils. J Mol Biol 193:115–125

Hodge AJ, Petruska JA (1963) Recent studies with the electron microscope on ordered aggregates of the tropocollagen macromolecule. In: Ramachandran GN (ed) Aspects of protein structure. Academic Press, London, pp 289–300

Hodges RS, Sodek J, Smillie LB, Jurasek L (1972) Tropomyosin: amino acid sequence and coiled-coil structure. Cold Spring Harb Symp Quant Biol 37:299–310

Huggins ML (1943) The structure of fibrous proteins. Chem Rev 32:195–218

Hulmes DJS, Miller A, Parry DAD, Piez KA, Woodhead-Galloway J (1973) Analysis of the primary structure of collagen for the origins of molecular packing. J Mol Biol 79:137–148

Huxley HE, Perutz M (1951) Polypeptide chains in frog sartorius muscle. Nature 167:1054

Kajava AV, Squire JM, Parry DAD (2006) Beta-structures in fibrous proteins. In: Kajava AV, Squire JM, Parry DAD (eds) Fibrous proteins: Amyloids, prions and beta proteins. Adv Protein Chem 73:1–15

Knupp C, Squire JM (2005) Molecular structure of network-forming collagens. In: Parry DAD, Squire JM (Eds) Fibrous proteins: Coiled-coils, collagen and elastomers. Adv Prot Chem 70:375–403

Kühlbrandt W (2014) Cryo-EM enters a new era. Elife 3:e03678

Kühnel K, Jarchau T, Wolf E, Schlichting I, Walter U, Wittinghofer A, Strelkov SV (2004) The VASP tetramerization domain is a right-handed coiled-coil based on a 15-residue repeat. *Proc Natl Acad Sci USA* 101:17027–17032

Lucas F, Rudall KM (1968) Extracellular fibrous proteins: the silks. In: Florkin M, EH Stotz (eds) Comprehensive Biochemistry, vol 26B. Elsevier, Amsterdam, pp 475–558

Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol 128:82–97

McMullan G, Faruqi AR, Henderson R, Guerrini N, Turchetta R, Jacobs A, van Hoften G (2009) Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. Ultramicroscopy 109:1144–1147

McMullan G, Faruqi AR, Clare D, Henderson R (2014) Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. Ultramicroscopy 147:156–163

Orgel JP, Miller A, Irving TC, Fischetti RF, Hammersley AP, Wess TJ (2001) The in situ supermolecular structure of type I collagen. Structure 9:1061–1069

Orgel JPRO, Persikov AV, Antipova O (2014) Variation in the helical structure of native collagen. PLoS One 9(2):e89519. doi:10.1371/journal.pone.0089519

Parry DAD (1988) The molecular and fibrillar structure of collagen and its relationship to the mechanical properties of connective tissue. Biophys Chem 28:195–209

Parry DAD, Elliott A (1967) The structure of a paracrystalline phase of poly-gamma-benzyl-L-glutamate in dimethylformamide. J Mol Biol 25:1–13

Parry DAD, Fraser RDB, Squire JM (2008) Fifty years of coiled-coils and α-helical bundles: a close relationship between sequence and structure. J Struct Biol 163:258–269

Paul D, Padwardhan A, Squire JM, Morris EP (2004) Single particle analysis of filamentous and highly elongated macromolecular assemblies. J Struct Biol 148:236–250

Paul DM, Squire JM, Morris EP (2010) A novel approach to the structural analysis of partially decorated actin based filaments. J Struct Biol 170:278–285

Pauling L, Corey RB (1951a) The structure of hair, muscle and related proteins. Proc Natl Acad Sci U S A 37:261–271

Pauling L, Corey RB (1951b) The structure of fibrous proteins of the collagen-gelatin group. Proc Natl Acad Sci U S A 37:272–281

Pauling L, Corey RB (1951c) Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. Proc Natl Acad Sci U S A 37:729–740

Pauling L, Corey RB (1953a) Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type. Nature 171:59–61

Pauling L, Corey RB (1953b) Two rippled-sheet configurations of polypeptide chains, and a note about the pleated sheets. Proc Natl Acad Sci U S A 39:253–256

Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A 37:205–211

Perutz MF (1949) An X-ray study of horse methemoglobin. Proc R Soc A 195:474–499

Perutz MF (1951) New X-ray evidence on the configuration of polypeptide chains. Nature 167:1053–1054

Rich A, Crick FHC (1961) The molecular structure of collagen. J Mol Biol 3:483–506

Rudall KM (1962) Silk and other cocoon proteins. In: Florkin M, Mason HS (eds) Comparative biochemistry, vol 4. Academic, New York, pp 397–433

Rudall KM (1965) Skeletal structure in insects. In: Aspects of insect biochemistry. Academic, London/New York, pp 83–92

Sanger F (1959) Chemistry of insulin. Science 129:1340–1344

Serpell LC. (2013) X-ray fibre diffraction. In: Encyclopedia for biophysics. Springer, Heidelberg. ISBN:978-3-642-16711-9

Squire JM (1981) The structural basis of muscular contraction. Plenum Press, New York

Squire JM, Elliott A (1972) Side-chain conformations in dry and liquid-crystalline forms of racemic poly-γ-benzyl glutamate. J Mol Biol 65:291–321

Squire JM, Vibert PJ (eds) (1987) Fibrous protein structure. Academic, New York/London

Stone D, Sodek J, Johnson P, Smillie LB (1975) In: Biro ENA (ed) Proceedings of the 9th FEBS Meeting, Proteins of Contractile Systems, vol 31. Akademiai Kiado, Budapest, and North-Holland Publishing, Amsterdam, pp 125–136

Subramaniam S, Kühlbrandt W, Henderson R (2016) CryoEM at IUCrJ: a new era. IUCRJ 3:3–7

van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A (2000) Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys 33:307–369

Wess TJ (2005) Collagen fibril form and function. In: Parry DAD, Squire JM (eds) Fibrous proteins: Coiled-coils, collagen and elastomers. Adv Protein Chem 70:341–374

Chapter 2
Coiled-Coil Design: Updated and Upgraded

Derek N. Woolfson

Contents

Abstract α-Helical coiled coils are ubiquitous protein-folding and protein-interaction domains in which two or more α-helical chains come together to form bundles. Through a combination of bioinformatics analysis of many thousands of natural coiled-coil sequences and structures, plus empirical protein engineering and design studies, there is now a deep understanding of the sequence-to-structure relationships for this class of protein architecture. This has led to considerable success in rational design and what might be termed *in biro de novo* design of simple coiled coils, which include homo- and hetero-meric parallel dimers, trimers and tetramers. In turn, these provide a toolkit for directing the assembly of both natural proteins and more complex designs in protein engineering, materials science and synthetic biology. Moving on, the increased and improved use of computational design is

D.N. Woolfson (✉)
School of Chemistry, University of Bristol, BS8 1TS Bristol, UK

School of Biochemistry, University of Bristol, BS8 1TD Bristol, UK

BrisSynBio, Life Sciences Building, University of Bristol, BS8 1TQ Bristol, UK
e-mail: d.n.woolfson@bristol.ac.uk

allowing access to coiled-coil structures that are rare or even not observed in nature, for example α-helical barrels, which comprise five or more α-helices and have central channels into which different functions may be ported. This chapter reviews all of these advances, outlining improvements in our knowledge of the fundamentals of coiled-coil folding and assembly, and highlighting new coiled coil-based materials and applications that this new understanding is opening up. Despite considerable progress, however, challenges remain in coiled-coil design, and the next decade promises to be as productive and exciting as the last.

Keywords Coiled coil • Computational design • *De novo* design • Peptide assembly • Protein design • Protein engineering

2.1   Scope of This Review

Arguably, the α-helical coiled coil is one of the most studied and the best understood of all protein structures (Lupas 1996a; Gruber and Lupas 2003; Woolfson et al. 2012). This review focuses on how our basic knowledge of coiled coils is advancing, and how this understanding is being translated into *de novo* coiled-coil design. The term *de novo design* can be defined as the process of generating completely new peptide and protein sequences that fold and assemble into prescribed 3D structures; this can be done rationally using heuristics, or rules of thumb (Woolfson et al. 2012; Regan et al. 2015), although increasingly it is achieved using computers to sample large swathes of sequence and structural space (Davey and Chica 2012; Feldmeier and Hocker 2013; Regan et al. 2015; Woolfson et al. 2015). It can be distinguished from *peptide / protein engineering* or *redesign*, which may be regarded as the iterative processes of altering usually natural sequences to effect predictable changes to protein structure, stability, and/or function (Magliery 2015). That said, with a large number of successful *de novo* coiled-coil designs now established, these two areas are merging: successful *de novo* coiled-coil designs are being used as templates onto which functions are being grafted (Mocny and Pecoraro 2015), and as building blocks for the construction of more-complex structures, large assemblies, materials and even systems (Bromley et al. 2008; Channon et al. 2008). This review attempts to cover all of these emerging areas. It builds on a previous review (Woolfson 2005), which outlined the basic concepts of coiled-coil design and catalogued *de novo* designs from the preceding 20 years or so.

Although this chapter develops themes from the 2005 article, the latter should be viewed as a basic introduction to the field of coiled-coil design. Because the field has developed considerably over the past 10 years, this new review has been structured differently. Firstly, what can be and has been achieved using established and relatively straightforward sequence-to-structure relationships in *classical coiled* coils is summarised. Secondly, there is a discussion of how the field is moving on in

terms of understanding coiled-coil sequences and structures more generally. This requires us to adjust and expand our traditional view of the sequence repeats and helix packing. Specifically, we have to consider multiple helix-helix interactions within coiled-coil structures. These lead to higher-order oligomer states and more complex coiled-coil assemblies. Thirdly, the development and use of computational methods, which are necessary to explore this larger design space, is discussed. Fourthly, there is a description of what is being done with both the classically-derived and more complex coiled coils, in terms of building elaborate peptide / protein-based architectures and materials, and also incorporating function into these assemblies. Finally, there is a short section on perspectives and where the field might be headed.

2.2   The Basics of Coiled-Coil Sequence and Structure

As the 2005 review covered many of the basic underlying concepts and foundational work in coiled-coil design (Woolfson 2005), this chapter will not dwell on these. Instead, it will discuss only old and new concepts central to the designs being reviewed herein.

However, for completeness and for the avoidance of doubt, coiled-coil units are defined as follows: coiled coils comprise two or more α-helices that wrap around each other to form supercoiled quaternary structures (Lupas 1996a; Lupas and Gruber 2005), Fig. 2.1a, b. As defined by Crick (Crick 1953b), the helix-helix interactions are directed and cemented by so-called knobs-into-holes (KIH) interactions, Fig. 2.1c, d. These are intimate interactions in which a sidechain (the knob) from one helix inserts into a diamond-shaped arrangement (the hole) projecting from another helix. It is generally accepted that helical assemblies must have contiguous stretches of KIH interactions in order to qualify as α-helical coiled-coil structures (Walshaw and Woolfson 2001b), otherwise they are simply globular α-helical domains or bundles, where different and less-intimate packing arrangements operate (Chothia et al. 1981; Walther et al. 1996). KIH-based structures relate back to underlying sequence repeats of hydrophobic (**h**) and polar (**p**) residues (Crick 1953b). Most commonly, these are heptad repeats, **hpphppp**, often denoted **abcdefg** (Lupas 1996a). Variations on this pattern are well known, and these have structural and functional consequences (Brown et al. 1996; Hicks et al. 1997, 2002; Gruber and Lupas 2003). Because the average 3.5-residue spacing of hydrophobic side chains in canonical heptad repeats closely matches the 3.6-residue repeat of the ideal α-helix, they result in amphipathic helices, Fig. 2.2a. In turn, and in what may be termed a *classical coiled coil*, two or more such helices assemble into helical bundles, Fig. 2.1a, b. However, because 3.5 and 3.6 are different, rather than packing like straight and rigid straws, the helices wrap or supercoil around each other.

A *classical coiled coil* is usually considered as one founded on a single, contiguous heptad pattern. To all intents and purposes, these have a single hydrophobic seam defined by residues at the **a** and **d** sites of the heptad repeat. Largely, this leads

Fig. 2.1 Coiled coils are
founded on knobs-into-
holes packing of α-helices.
(**a** & **b**) Orthogonal
projections of the X-ray
crystal structure of an
α-helical coiled-coil dimer
(2ZTA, (O'Shea et al.
1991)). (**c** & **d**) Orthogonal
projections of an ***a***-knob
projecting into ***d'g'a'd'***-
hole from the same
structure as panels (**a** & **b**).
Key: ***a** red*, ***d** green*, ***g***
violet (Protein-structure
images were generated in
PyMOL (www.pymol.org))
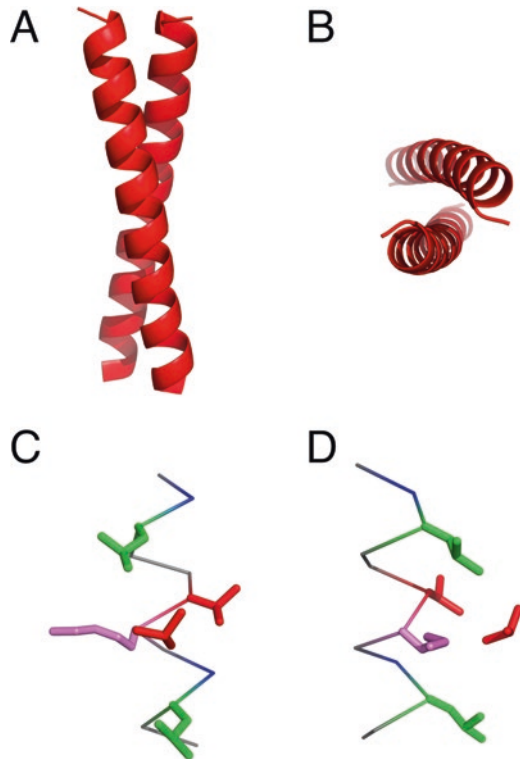
to only dimeric, trimeric and tetrameric coiled-coil assemblies. However, as introduced and discussed below, our current understanding of coiled coils, and the desire to delve deeper into the structural space possible for coiled coils, requires us to move on from this classical treatment and to consider *more-complex*, or *extended coiled-coil interfaces* (Walshaw and Woolfson 2001a, 2003; Moutevelis and Woolfson 2009; Woolfson et al. 2012).

2.3 Computational Tools for Analysing Coiled-Coil Sequences and Structures

There is now a raft of well-established and reliable web-based bioinformatics tools available:

1. For examining protein sequences to locate coiled-coil regions (Lupas et al. 1991; Lupas 1996b; Delorenzi and Speed 2002)
2. To predict coiled-coil oligomer state (McDonnell et al. 2006; Rackham et al. 2010; Armstrong et al. 2011; Trigg et al. 2011; Vincent et al. 2013; Ramisch et al. 2015)
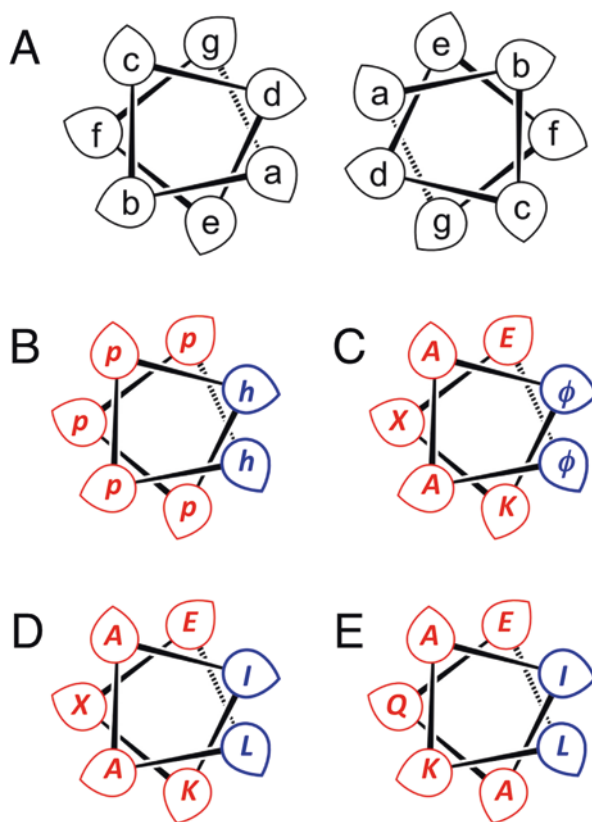
Fig. 2.2 Coiled-coil sequence repeats visualised by helical-wheel diagrams. (**a**) The heptad repeat, **abcdefg**, spun out onto helical wheels for a classical dimeric coiled coil. The central Cα atom of each residue is represented by "teardrop" with the point directed approximately towards the Cβ atom. (**b**) The repeat of hydrophobic (**h**) and polar (**p**) residues for a classical coiled coil. (**c**) Helical wheel giving the heptad background upon which the CC-series of *de novo* coiled coils were built (Zaccai et al. 2011; Fletcher et al. 2012). (**d & e**) Helical wheels showing the sequence relationship between CC-Tet and CC-Hex (Zaccai et al. 2011). For all of these wheels, the number of residues per turn is 3.5, and not 3.6 for a standard α-helix, *i.e.* the helical wheels effectively uncoil the super-helical coil

3. For analysing X-ray crystal and NMR structures of coiled-coil domains that have been deposited in the RCSB Protein Data Bank (PDB; (Rose et al. 2015))

For example, the program SOCKET (Walshaw and Woolfson 2001b) identifies KIH interactions in PDB coordinate files and, along with this, TWISTER can determine geometrical parameters for coiled-coil structures (Strelkov and Burkhard 2002). Moreover, outputs from these programs are being collated into useful relational databases and schemes to visual and interrogate coiled-coil sequences and structures rapidly and *en masse* (Moutevelis and Woolfson 2009; Testa et al. 2009). Specifically, the database (Testa et al. 2009) now has >2500 non-redundant

entries, representing a significant resource for the coiled-coil and wider community. CC+ can be searched using a number of easy-to-use pull-down menus within four over-arching tabs for keywords, structures, sequences and interactions. In this way, a user can readily and quickly create a subset of coiled-coil structures for further manual inspection or computational analysis. Indeed, CC+ is being used in this way by others to search for structure-function and evolutionary relationships of natural coiled coils (Surkont et al. 2015), and to benchmark the performance of coiled-coil modelling and design algorithms (Ramisch et al. 2015).

2.4 Straightforward Rules for Coiled-Coil Prediction and Design

The link between sequence and structure makes coiled coils ideal targets for rational design of peptides and proteins. Indeed, such is the depth of our understanding of classical coiled coils that sequences that adopt parallel dimers, trimers and tetramers can simply be written down (Harbury et al. 1993; Fletcher et al. 2012). This follows from seminal work conducted by Harbury, Alber and Kim on hydrophobic-core variants of the leucine-zipper region of a yeast transcriptional activator, GCN4-p1 (O'Shea et al. 1991), and work that followed (Harbury et al. 1994; Woolfson and Alber 1995; Gonzalez et al. 1996; Woolfson 2005). This furnished us with straightforward sequence-to-structure heuristics for classical coiled coils. In GCN4-p1, which forms a parallel dimer, all *d* sites are leucine (Leu, L), and, although more varied, the canonical residue at *a* is valine (Val, V). Harbury's work shows that different combinations of Leu and isoleucine (Ile, I; which is β-branched like Val) at the *a* and *d* sites, lead to different oligomer states being formed. For example, the combination of *a* = *Ile* and *d* = *Leu* promotes parallel dimers, *a* = *d* = *Ile* specifies trimers, and *a* = *Leu* plus *d* = *Ile* directs tetramers. Although we have learnt over the past decade that things are a little more complicated, these heuristics largely hold up to mutational studies of natural proteins, and completely *de novo* designs (Woolfson 2005; Woolfson et al. 2012). Thus, they allow what might be termed *in biro* design of coiled coils; indeed, this is so well embedded in peptide and protein science that specialist and non-specialist labs can turn out designs of coiled coils and translate these into experiments with ease and, more importantly, with predictable outcomes. In short, *in biro* coiled-coil design is now *de rigueur*.

The *Harbury relationships* work because they link sequence and final structure, in this case the oligomer state of the coiled coil. The variants can be rationalised in structural terms that are fully consistent with Crick's original KIH postulate, Fig. 2.1c, d. Because of helical geometry, the Cα-Cβ bond vector of a sidechain projects out in a defined way from the helical backbone, Fig. 2.3a. As a result, when two helices come together in a parallel dimer, sidechains at *a* and *d* sites (the knobs) project differently towards the constellation of sidechains on the partnering helix that form their recipient holes, as illustrated in Figs. 2.1c, d and 2.3b. For the *d* sites,

the resulting packing is *perpendicular—i.e.*, the knob points directly into the hole— which is favoured by the non-β-branched, hydrophobic alkyl sidechain of Leu. Indeed, this is borne out by analysis of all non-redundant parallel coiled-coil dimers in the PDB, >50 % of the *d* sites are Leu (Testa et al. 2009). By contrast, the Cα-Cβ bond vectors of the *a* knobs are *parallel* to each other. This type of packing is more accommodating of different sidechain shapes, hence the acceptance of, and indeed preference for β-branched Val and Ile residues. Adding helices to form parallel trimers and tetramers results in different KIH geometries, Fig. 2.3c, d. This can be thought of as the helices rotating as a consequence of the hydrophobic moment of the helix, which falls between the hydrophobic *a* and *d* sites, being required to point towards the centre of the assembly. As a result, the KIH geometries of how the *a* and *d* knobs point into their respective holes changes. From dimer to trimer, the *a* and *d* KIH geometries move from the extremes of parallel and perpendicular packing, respectively, to being more similar, and in between these two states, in so-called *acute packing*. This then explains why in trimers the amino-acid preferences at the two sites are more similar, which is again borne out on inspection of natural sequences and structures. Adding a further helix to make a tetramer rotates the helices relative to one another again, and with it the KIH geometries change. In this case, the *a* and *d* side chains move through intermediate acute packing to perpendicular and parallel packing, respectively, Fig. 2.3d. Thus, the packing geometries are swapped from those seen with parallel dimers, explaining the residue preferences swap in the GCN4 mutants. Though there are fewer sequences in the CC+ databases for tetramers compared with dimers and trimers, this broad-brush rule for *a = Leu* plus *d = Ile* appears to hold in natural tetramers too.

A final point to note here is that whereas the *a* and *d* KIH interactions are complementary in dimers, they daisy chain in trimers and tetramers; that is, a knob from helix A interfaces with helix B, and the corresponding knob from helix B interfaces with helix C and so on, Fig. 2.3.
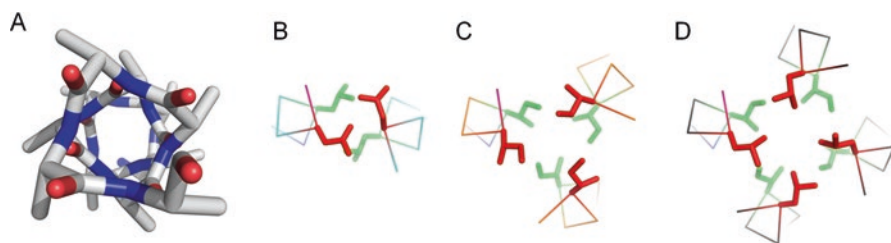
Fig. 2.3 Classical coiled-coil dimers, trimers and tetramers differ in their knobs-into-holes packing. (**a**): The projection of the Cα—Cβ vectors (*grey sticks*) of sidechains from a standard α-helix. (**b–d**): Similar diagrams for dimeric, trimeric and tetrameric coiled coils, respectively. These diagrams, which focus on a single heptad repeat and highlight the *a* (*red*) and *d* (*green*) sites only, show that although the sidechains project in the same ways from each helix, they are directed differently towards neighbouring helices in the three coiled-coil oligomers. This is discussed in detail in the text (Protein-structure images were generated in PyMOL (www.pymol.org))

2.5 A Heptad of Completely de novo Helical Assemblies

2.5.1 Rationally Designed Dimers, Trimers and Tetramers

As elegant as they are, the studies of Harbury *et al.* were performed in the background of a natural coiled coil, namely GCN4-p1, and this sequence may be complicated by biological function and evolutionary pressure. For two reasons Harbury's relationships have been tested in a completely *de novo* background (Fletcher et al. 2012). First, it was important to know if the rules were context-independent or not, *i.e.* if they could be translated to other systems; second, the aim was to generate a *basis set* or *toolkit of coiled-coil modules* for further protein-engineering and synthetic-biology projects. The rationale for the latter was that a set of well-characterised and context-independent coiled-coil modules would have considerable applications for directing protein assembly, and that these could be orthogonal to natural protein-protein-interaction motifs (Bromley et al. 2008; Channon et al. 2008).

To do this, a general background for parallel coiled-coil bundles, *i.e.* EΦAAΦKX, has been designed which maps on to a ***gabcdef*** heptad repeat, Fig. 2.2c. In these sequences, combinations of Ile and Leu are used at the ***a*** and ***d*** sites (denoted Φ) to direct dimer, trimer and tetramer formation as informed by Harbury (Harbury et al. 1993); glutamate (E) and lysine (K) placed at ***g*** and ***e*** sites to encourage inter-helix salt bridges flanking the hydrophobic core; and the ***b, c*** and ***f*** sites are made *vanilla*, with alanine (A) at ***b*** and ***c*** to encourage helicity, and glutamine (Q) and/or lysine at ***f*** (denoted X) for water solubility. Different chromophore and mass tags are used to aid characterisation and, finally, the peptides are four heptads long for stability.

Synthetic peptides for the initial design iterations (CC-pIL, CC-pII and CC-pLI) are all water-soluble, fully α-helical and highly thermally stable (Fletcher et al. 2012). Whilst solution-phase measurements and X-ray crystallography confirm CC-pII and CC-pLI as parallel trimer and tetramer, respectively, CC-pIL is a trimer and not a dimer. Though surprising in the light of Harbury's data, this fits with the broader body work on sequences of the type ***a*** = Ile/Val plus ***d*** = Leu, which can adopt various oligomer states (Ogihara et al. 1997; Ghirlanda et al. 2002; Oshaben et al. 2012). As it is well known that a single asparagine (Asn, N) at ***a*** in natural leucine-zipper and related sequences help specify dimers (Woolfson and Alber 1995; Gonzalez et al. 1996; Lumb and Kim 1998), the centre-most ***a*** in CC-pIL has been mutated from Ile→Asn. This gives the peptide of reduced thermal stability, though it is still a fully folded parallel dimer in solution and by X-ray crystallography. An analogous insertion of Asn at a ***d*** site of the CC-pII sequence, yields a peptide with reduced thermal stability, but remains a parallel trimer state, which is fully consistent with foregoing bioinformatics and experimental data (Hartmann et al. 2009). The net result is a set of fully characterised coiled-coil oligomers—renamed as CC-Di, CC-Tri and CC-Tet—complete with high-resolution X-ray crystal structures, Fig. 2.4a–c.
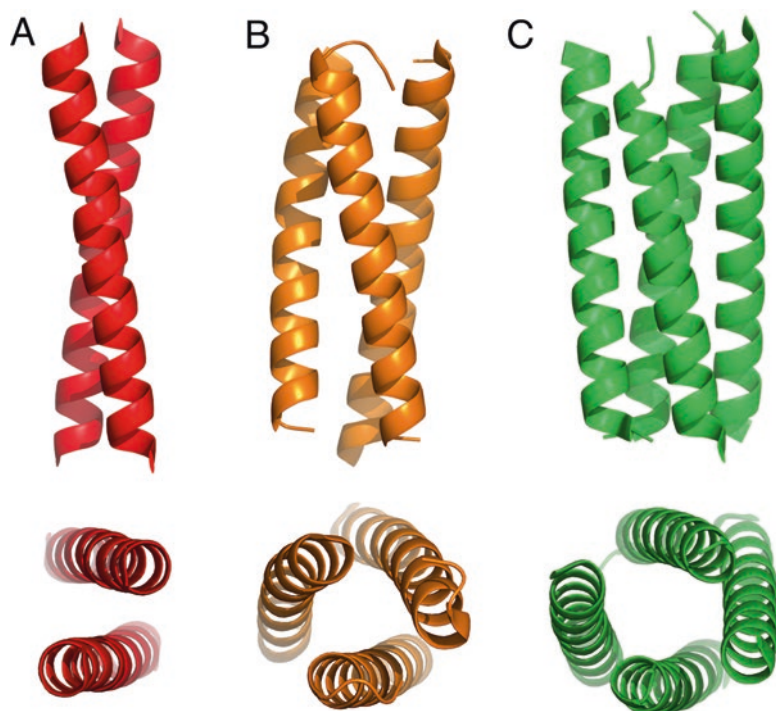
Fig. 2.4 A basis set of de novo coiled coils: dimer, trimer and tetramer. (**a–c**) Orthogonal projections of the X-ray crystal structures of CC-Di, CC-Tri and CC-Tet, respectively. The PDB codes for these are 4DZM, 4DZI and 3R4A, respectively (Zaccai et al. 2011; Fletcher et al. 2012) (Protein-structure images were generated in PyMOL (www.pymol.org))

2.5.2 Expanding de novo Coiled Coils Past Tetramer

Inspection of the CC+ database reveals that the vast majority of natural coiled-coil structures are for dimers, trimers and tetramers, Fig. 2.5. The emphasis and success of the design field on these oligomers states almost certainly reflects this bias in the natural coiled coils observed to date. That said, the door to larger oligomer states is ajar, and has been for a while. There are several pentamers, which include natural and engineered peptides and water-soluble and membrane-spanning proteins (Malashkevich et al. 1996; Liu et al. 2006c; Eshaghi et al. 2006; Payandeh and Pai 2006; Vostrikov et al. 2013; Sastri et al. 2014), one variant of GCN4-p1 forms a heptamer, albeit an unusual spiralling structure (Liu et al. 2006b), there are pore-like natural structures for a membrane-spanning octamer (Dong et al. 2006), a decamer (Sun et al. 2014), and a dodecamer (Koronakis et al. 2000). The interest in pursuing engineering and design of these larger assemblies is twofold. First, it would be interesting to know the limits of coiled-coil assemblies and, second, pentamers and above either have, or are predicted to have contiguous, open-ended,
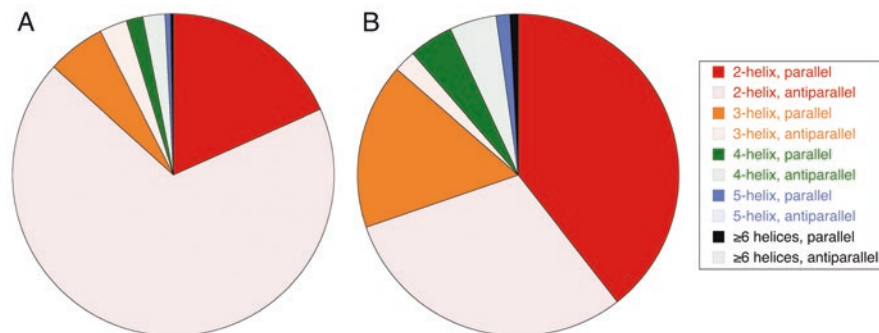
Fig. 2.5 Biased distribution of coiled-coil oligomer states and topologies in the PDB and CC+ database. (**a**) All coiled coil-containing proteins. (**b**) Only intermolecular coiled-coil motifs. The data were culled from the CC+ database of August 2015 (Testa et al. 2009) using the following settings: 70 % sequence redundancy, and canonical heptad repeats of >11 residues long. The raw counts of structures summarised in these plots are: parallel dimers (**a**, 403; **b**, 305); antiparallel dimers (**a**, 1503; **b**, 233); parallel trimers (**a**, 128; **b**, 128); mixed/antiparallel trimers (**a**, 65; **b**, 19); parallel tetramers (**a**, 33; **b**, 33); mixed/antiparallel tetramers (**a**, 50; **b**, 36); parallel pentamers (**a**, 12; **b**, 12); mixed/antiparallel pentamers (**a**, 0; **b**, 0); parallel structures with ≥6 helices (**a**, 5; **b**, 1); mixed/antiparallel with ≥6 helices (**a**, 5; **b**, 1)

central channels or pores; *i.e*, they are *α-helical barrels*. Indeed, the observed naturally-occurring pentamers and the octamer, decamer and dodecamer bind or transport other biomolecules. However, for these structures there are not enough examples or data to inform designs as described above. However, a route into the design of α-helical barrels has been discovered serendipitously as follows.

A permutant of CC-Tet has been made with the residues at the *b* and *e* positions exchanged; *i.e.*, the *g→f* heptad repeat was changed from ELAAIKX to ELKAIAX, Fig. 2.2d, e. The resulting peptide is fully α-helical and highly thermally stable. Surprisingly, the solution-phase oligomer state is hexamer, which is confirmed by a high-resolution X-ray crystal structure that reveals a parallel 6-helix supercoiled bundle with regular KIH interactions (Zaccai et al. 2011). This novel structure also has a central ~6 Å channel running completely through its centre. Furthermore, the peptide CC-Hex is somewhat, though not completely, robust to mutations within the lumen of the structure (Burton et al. 2013; Burgess et al. 2015; Thomas et al. 2016). Until that point, there was only one other coiled-coil-based hexamer in the PDB, which is an antiparallel structure buttressed by other helices (Tanaka et al. 2007).

With a view (1) to engineering more-robust versions of CC-Hex, and (2) to designing rare oligomers other than hexamer predictively, a computational design programme has been initiated to generate coiled-coil pentamers, hexamers and heptamers. This has delivered designs and X-ray crystal structures for CC-Pent, CC-Hex2, CC-Hex3, and CC-Hept, Fig. 2.6a–c (Thomson et al. 2014). In parallel, the Baker group has described the parametric computational design of hyperthermostable coiled-coil trimers, tetramers and a pentamer, Fig. 2.6d (Huang et al. 2014). More recently, André and colleagues have reported the design of another
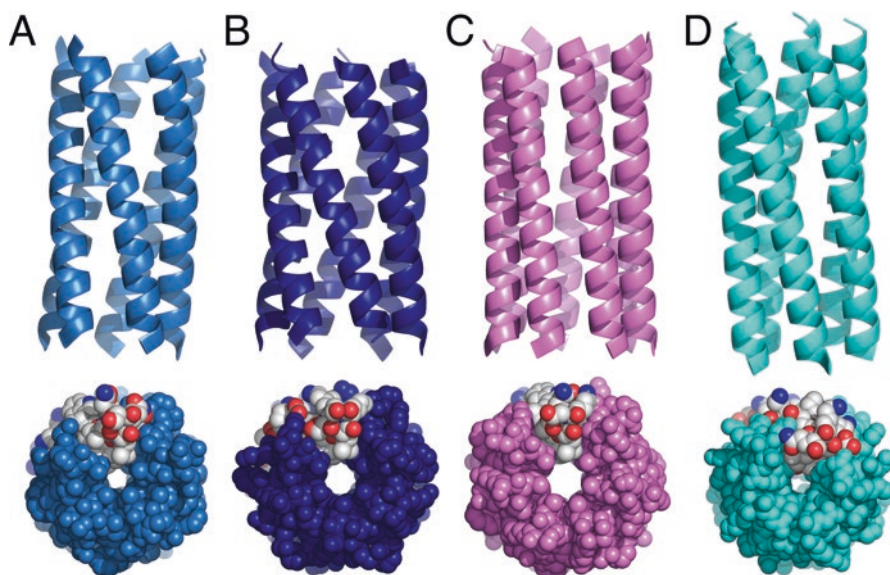
Fig. 2.6 Coiled coils beyond tetramer are possible by design. (**a–c**) Orthogonal projections for the X-ray crystal structures of CC-Pent, CC-Hex2 and CC-Hept, respectively. The PDB codes for these are 4PN8, 4PN9 and 4PNA, respectively (Thomson et al. 2014). (**d**) A computationally-designed pentamer from the Baker group (4UOT, (Huang et al. 2014)). *Top*: Ribbon diagrams tracing the polypeptide backbones. *Bottom*: Space-filling representations with one chain in each structure coloured by atom type (Images were generated in PyMOL (www.pymol.org))

pentamer, which switches to a 6-helix bundle upon change of pH (Lizatovic et al. 2016). Details for these studies are given below.

2.5.3   Structural Rationale for Higher-Order Oligomerization

The core-packing arguments learned from classical dimers to tetramers (Harbury et al. 1993; Fletcher et al. 2012) extend to explain the higher-order coiled coils, such as CC-Hex. Above tetramer, apart from small changes with increasing oligomer state, KIH geometries for the traditional *a* and *d* sites effectively remain fixed at perpendicular and parallel, respectively. As a result, the *largermers*, *i.e.* CC-Pent, CC-Hex and CC-Hept, appear best specified by *a* = Leu plus *d* = Ile (Thomson et al. 2014; Huang et al. 2014), although variations on this are possible. This is also largely the case for the small number of known natural pentamers, such as COMP (Malashkevich et al. 1996; Testa et al. 2009). Therefore, the controlling influence that selects between tetramers at least through to heptamers must reside elsewhere.

The addition of helices in classical dimers through tetramers leads to the increased inclusion of peripheral residues within the hydrophobic core of the coiled-coil assembly; notably, residues at *g* become knobs. These are referred to as Type I

Fig. 2.7 Variations on the classical heptad repeat – different repeats lead to different ranges of oligomer states. Helical wheels for Types N, I, II and III coiled-coil repeats. *Blue* and *green* teardrops represent knob-forming and nominally hydrophobic residues; *red* teardrops represent the remaining and usually polar residues; and the *dark green* angles show the angle subtended between the two resulting hydrophobic seams of Type I – III repeats. Type N repeats are found predominantly in dimers. Based on the offset angles shown, trimers and tetramers are founded on Type I repeats; and higher-order structures on Type II (tetramer – heptamer) and Type III (octomer and above) repeats, although these are rules of thumb rather than hard-and-fast rules. The overall hydrophobic/polar sequence patterns are given below each helical wheel
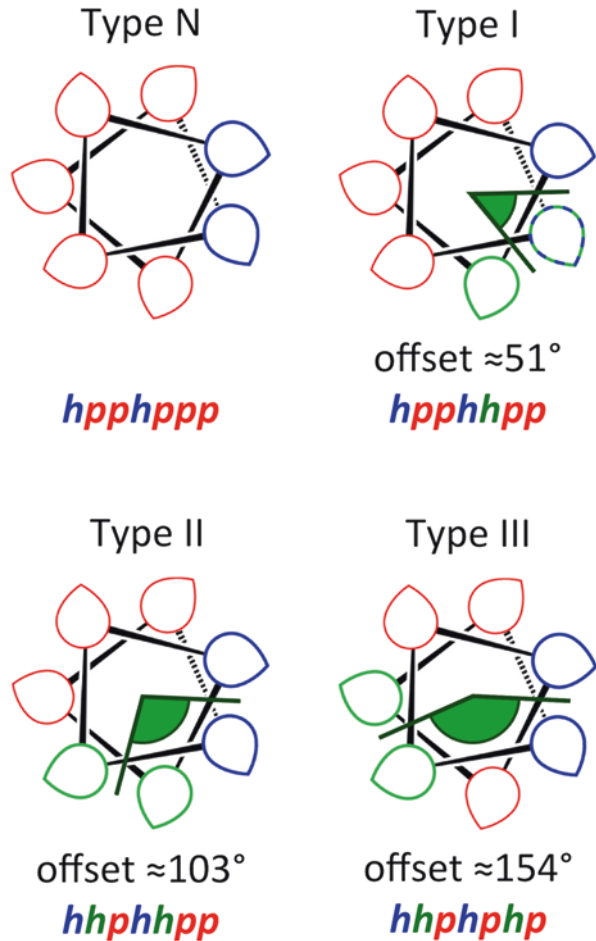
interfaces, and are distinct from classical Type N interfaces, Fig. 2.7a, b, (Woolfson et al. 2012; Walshaw and Woolfson 2003). Increasing the helix-helix contacts in pentamers and above leads to Type II interfaces, in which sidechains at both *g* and *e* sites are involved in KIH interactions, Fig. 2.7c. For completeness, Type III interfaces are the final variations in this scheme, and these lead to barrels of the type observed in the 12-helix barrel of TolC, Fig. 2.7d (Koronakis et al. 2000; Walshaw and Woolfson 2001a).

Another way to consider this is that coiled-coil pentamers and above comprise two superposed dimer sequences with the two traditional *a* and *d* hydrophobic seams falling at *e* and *a*, and *d* and *g* positions, respectively, Fig. 2.7c. More specifically, and with the traditional heptad nomenclature as a reference, the *a* and *d* sites are predominantly occupied by Leu and Ile residues, respectively, and the flanking *e* and *g* sites are occupied by residues that effectively fine-tune oligomer-state selection. For the smaller oligomer, CC-Pent, the flanking residues are large Ile and Glu

sidechains. In the hexameric variants, these sites are occupied by Glu or Ser at *g* plus Ala at *e*. For the heptamer, both residues are Ala. In short, the smaller the volume of the sidechains at *e* and *g*, the more the helix-helix-helix packing angle can open up, and the larger the barrel (Thomson et al. 2014). Baker's designed pentamer conforms to these rules of thumb, with all of the *a* sites Leu, the *d* sites combinations of Ile, Gln and Leu, and the flanking *e* and *g* sites combinations of Leu, Met, Ile and Glu (Huang et al. 2014), although the single β-branched residue at *d* is intriguing.

This section promised a heptad of helical folds, but so far has delivered only six (dimer through heptamer). Recently, we have reported a series of highly α-helical, but monomeric peptides (Baker et al. 2015). These indicate that whilst the so-called macrodipole (Hol et al. 1978) of the α-helix may exist, it is outweighed by local electrostatic effects and does not contribute noticeably to helix stability. Though this may seem somewhat of a stretch in this review, it is pertinent to coiled-coil folding and assembly. It is often heard said that parallel arrangements of the helices in coiled coils should be disfavoured over antiparallel forms because of (assumed) unfavourable interactions between adjacent helix macrodipoles in the former. However, if these macrodipoles are masked, or dominated by other forces, and do not contribute significantly to stability, then that debate should be ended.

This does, of course, raise the question: *why are there so many antiparallel pairs of helices in the CC+ database?* This seems best explained in that these tend to be parts of single-chain coiled coil-containing proteins, and as helix-loop-helix topologies. This can be seen by comparing panels A and B of Fig. 2.5. In the former, all coiled-coil structures are counted, including intramolecular coiled coils. Many of the 2-helix structures in this category will be helix-loop-helix structures. By contrast, panel B is based only on the intermolecular coiled coils, and therefore cannot be biased in this way. Indeed, this second panel indicates a much lower proportion of antiparallel dimers in particular, which we argue represent a reduced tendency of intermolecular coiled coils to form antiparallel arrangements.

2.6 Parametric and Computational Coiled-Coil Design

2.6.1 Background: Computational Methods Old and New

The application of computational methods has a long and strong history both in the analysis of natural coiled-coil sequences and structures (*vide supra*) (Lupas 1996b), and the design of *de novo* coiled coils (Harbury et al. 1998). Recent advances that build on these foundations are having a considerable impact in coiled-coil design, and it is predicted that they will increasingly do so; indeed, improved confidence, predictability and ambition in the design of coiled-coil peptides and proteins will likely only come with the increased development and application of computational methods to these problems.

A key aspect of coiled coils is that they are inherently parameterizable; indeed, they were predicted parametrically by Crick (1953a, b) rather than being discovered empirically through X-ray crystallography. Moreover, the number of parameters needed to describe them is small, at least for regular coiled-coil structures. This has led to a number of successful and useful computational resources to model coiled-coil assemblies (Offer et al. 2002; Harbury et al. 1995; Offer and Sessions 1995; Grigoryan and DeGrado 2011; Wood et al. 2014; Ramisch et al. 2015), and, more recently, to design them (Plecs et al. 2004; Harbury et al. 1998; Grigoryan et al. 2011; Huang et al. 2014; Thomson et al. 2014; Lizatovic et al. 2016). In essence, all of these build on Crick's parameterization of the coiled coil: they use the Crick equations to generate backbones, and then add various methods to (i) add sidechains to the models, (ii) assess some *in silico* energy of the generated state, and (iii) search through combinations of structural parameters and sequence space to find solutions for the design target to test experimentally.

Of particular note are two web-based and user-friendly tools—CCCP and CCBuilder (Grigoryan and DeGrado 2011; Wood et al. 2014)—which make coiled-coil modelling and design accessible to all, and expand possibilities for designing coiled-coil structures both to mimic nature and to move into the *dark matter of protein structures* (Taylor et al. 2009).

2.6.2   Parametric Coiled-Coil Designs Achieved to Date

Parametric structure-based designs have been realised before now for coiled-coil proteins, including right-handed structures that incorporate non-natural amino acids, to satisfy unusual packing in the interior of the helical assembly (Harbury et al. 1998; Plecs et al. 2004). However, CCCP, CCBuilder and methods developed in the Baker lab (Grigoryan and DeGrado 2011; Wood et al. 2014; Huang et al. 2014) now allow structure-based designs of coiled coils and helical bundles to be tackled more generally. A key feature of all of these methods is that they implicitly incorporate backbone variations, which has always been a challenge in computational protein design (Harbury et al. 1995; Harbury et al. 1998; Grigoryan and DeGrado 2011; Huang et al. 2014; Wood et al. 2014; Ramisch et al. 2015).

Grigoryan, DeGrado and co-workers have used CCCP to generate helices predicted to make barrel-like assemblies that bind carbon nanotubes (Grigoryan and DeGrado 2011). This opens potential applications requiring solubilised carbon nanotubes. Recently, the same group has applied their approach to design functional membrane-spanning helical bundles called "Rocker" (Joh et al. 2014). This is a zinc-binding four-helix bundle. The idea is that zinc ions presented to one side of the membrane are transported to the other side *via* exchange between two zinc-binding sites, and that this is facilitated both by the rocking of the structure and by the transport of protons in the opposite direction.

Baker's group has achieved computational parametric designs for hyperstable coiled coils (Huang et al. 2014). These water-soluble designs are for single-chain,

3- and 4-helix bundles, which have both parallel and antiparallel helix-helix contacts; and also a parallel 5-helix bundle, Fig. 2.6d. A key and elegant aspect of these designs is the focus on layers of hydrophobic residues that define the hydrophobic core, rather than sequence-based repeats *per se*. By considering 2-, 3- and 5-layer structures, which correspond to 7-, 11- and 18-residue sequence repeats (Hicks et al. 2002; Gruber and Lupas 2003), the X-ray crystal structures of the designs reveal a left-handed supercoil for the pentamer (4UOT), a four-helix bundle with a slight right-handed twist (4UOS), and a three-helix assembly with straight helices (4TQL) (Huang et al. 2014). In an elegant study, Baker's group have used Rosetta, supplemented with new approaches to generate hydrogen-bonded networks of side-chains *in silico*, to design and deliver a series of two-layer helical bundles incorporating intricate networks of buried polar residues (Boyken et al. 2016). Also using Rosetta-based methods, André and co-workers have implemented parametric coiled-coil modelling and scoring to design a peptide that forms a pentamer at pH 8, which surprisingly switches to a collapsed two-layer six-helix bundle (a trimer of antiparallel dimers) at pH 6 (Lizatovic et al. 2016). This last study, in particular, illustrates the complexity of the free-energy landscape for coiled-coil assemblies, and the resulting plasticity of these structures. Thus, it highlights the difficulties in the *de novo* design of coiled coils predictively and reliably.

Finally, and as introduced above, parallel pentameric, hexameric and heptameric α-helical barrels (CC-Pent, CC-Hex and CC-Hept) have all been built (Thomson et al. 2014). The new designs are based on the realization that successive helical interfaces in cyclic coiled-coil structures can be approximated as heterodimeric faces encoded within the same peptide chain, *i.e.* Type II interfaces, Fig. 2.7c. For these designs, potential sequences for the two halves of the heterodimeric interfaces are initially selected from millions of sequence combinations using a scoring algorithm developed by Keating (Fong et al. 2004). The selected sequences are then modelled as coiled-coil bundles and barrels ranging from tetramer through to octamer. The BUDE-based scoring function of CCBuilder is then used to predict the preferred oligomer state. A number of the final sequences are taken through to experimental characterisation in solution and by X-ray protein crystallography. For more than half of the peptides tested, the observed and predicted oligomer states match.

2.7   Improving Heterospecificity in Coiled-Coil Design

Although, *de novo* designs for heterodimeric coiled coils have been with us for some time (O'Shea et al. 1993; Moll et al. 2001; Litowski and Hodges 2002), there has been considerable progress over the past decade in defining and improving these systems. These principles for design and assembly will not be discussed further other than to note that early examples were mostly specified using oppositely charged residues at complementary *g* and *e* sites of the heptad repeat (Woolfson 2005), and this basic principles remains a key aspect of new designs. Recent

attention has switched to designing or selecting sets of mutually exclusive (orthogonal) heterodimer pairs (Bromley et al. 2009), and controlling specificity and stability in these systems, with Keating and colleagues leading the way here (Thompson et al. 2012; Potapov et al. 2015).

Continuing the theme of generating and populating peptide toolkits for protein engineering and synthetic biology (Bromley et al. 2008; Fletcher et al. 2012), several groups have constructed sets of heterodimeric coiled coils. For example, the aforementioned rules of thumb for coiled-coil dimers with computational methods to deliver orthogonal heterodimers (Bromley et al. 2009) have been combined, and variants made of what is called the CC-Di-AB system, with binding affinities in the μM – nM range (Thomas et al. 2013). Building on foregoing work to make a *coiled-coil interactome* (Reinke et al. 2010), Lim and Keating have introduced SYNZIPs, increasing the number of orthogonal heterodimers in one set to an impressive 27 pairs (Thompson et al. 2012). Some of these pairs have been used in synthetic biology to control gene activation in a model system. Other contributions from Keating's group include demonstrating that non-interfacial, surface-exposed residues influence binding anti-bZIP peptides designed to natural targets: specifically, that increasing the helix propensity at these sites improves binding (Kaplan et al. 2014); and, notably, the considerable advance of using knowledge-based computational methods to predict heterodimer stability (Potapov et al. 2015). This method has clear applications not only for generating highly specified coiled-coil pairings, but also in developing peptides to target and disrupt natural protein-protein interfaces. Additional sets of heterodimers have been contributed by Jerala (Gradisar and Jerala 2011), Aili (Aronsson et al. 2015), and Mason (Crooks et al. 2016).

It remains to be seen how orthogonal all of these sets are to each other. Nonetheless, they provide valuable additions to peptide toolkits for applications in protein engineering and synthetic biology. Specifically, the rigorous characterizations of some of these peptide designs, which has delivered internal orthogonality, experimentally determined K_D values, and in some cases X-ray crystal structures for a designed heterodimers (Reinke et al. 2010; Sharp et al. 2012). All of these attributes will improve the predictability of using the toolkits and the development of future designs.

Compared to this progress on heterodimers, new heterospecific systems for larger oligomers have progressed more modestly, which is perhaps understandable given the increased demands on design. As of 2005, there were designs for heterotrimers (Nautiyal et al. 1995; Nautiyal and Alber 1999; Kashiwada et al. 2000; Kiyokawa et al. 2004; Schnarr and Kennan 2002) and a heterotetramer (Fairman et al. 1996). Since then, Kennan has used polar contacts in place of previously used complementary hydrophobic sidechains (Schnarr and Kennan 2002) to make another heterotrimeric system (Diss and Kennan 2008); Fairman has extended his studies of the Lac repressor (Fairman et al. 1996) to a GCN4-p1 background, to give an alternative A_2B_2 design (Root et al. 2009); Tanaka's group have applied their IZ peptides to switch protein-complex structure and function (Mizuno et al. 2009); and a heterohexamer based on CC-Hex has been delivered (Zaccai et al. 2011).

2.8   Adding Antiparallel Coiled Coils to the Mix

The elephants in the room of coiled-coil structure and design are assemblies with antiparallel, or mixed arrangements of helices. This is apparent from Fig. 2.5, which indicates that there are many natural antiparallel structures. However, emphasis in the design community, at least as judged from publications, has been for small oligomers with parallel arrangements of helices. As noted in the 2005 review (Woolfson 2005) and Oakley's earlier review (Oakley and Hollenbeck 2001), there are notable exceptions, with successful designs of antiparallel dimers from Hodges, Chaiken, Regan, Oakley and from Woolfson (Monera et al. 1993, 1994; Myszka and Chaiken 1994; Ghosh et al. 2000; McClain et al. 2001; Gurnon et al. 2003; Pandya et al. 2004). Nonetheless, with the exception of the Coiled Serine 1cos, which came much earlier and is a 3-helix structure (Lovejoy et al. 1993), no *de novo* antiparallel coiled-coil structures had been fully verified by X-ray crystal structures.

There are signs that this is about to change. Over the past decade, there has been considerable progress in our fundamental understanding of antiparallel arrangements of helices in coiled coils, and in our abilities to model these structures (Apgar et al. 2008; Grigoryan and DeGrado 2011; Wood et al. 2014). Sequence-to-structure relationships in antiparallel coiled coils have been probed using bioinformatics for some time (Walshaw and Woolfson 2001b), using experimental approaches more recently (Hadley and Gellman 2006), and with combinations of the two (Hadley et al. 2008; Steinkruger et al. 2012a). Notably, Gellman has applied his backbone thioester exchange (BTE) method to determine energies of interaction between residue pairs or groups of three across synthetic, designed antiparallel interfaces (Hadley and Gellman 2006; Hadley et al. 2008; Steinkruger et al. 2012a). His group has also extended these studies to models for parallel dimers (Steinkruger et al. 2012b).

Negron and Keating report what should be considered the current state-of-the-art in antiparallel coiled-coil design (Negron and Keating 2014). Their paper describes the computational design and solution-phase experimental validation of *de novo* and, importantly, orthogonal antiparallel homodimers, using the group's CLASSY and DFIRE computational methods (Negron and Keating 2013, 2014). In addition, the paper puts the work in a broader context, and impressively so, by conducting deep computational analyses of the designs using aforementioned computational methods such as Rosetta, CCBuilder and LOGICOIL.

2.9   Building with Coiled-Coil Modules: Protein Origami, Synthetic Biology and Materials

The availability of defined and well-characterised coiled-coil building blocks, and indeed peptide and protein units in general, paves the way for the reliable construction of more-complex assemblies and materials (Sinclair 2013; Kocar et al. 2015;

Quinlan et al. 2015). In the language of supramolecular assembly, these might be referred to as *tectons*, and in synthetic biology as *modules* or *parts* (Bromley et al. 2008; Channon et al. 2008). As outlined below, module-based assembly brings advantages of speed, more predictable outcomes, and the exchangeability of parts to the design and construction process. Roughly in order of increasing complexity, the following have been achieved using coiled coils as tectons.

CC-Tri to direct the assembly of difficult-to-fold bacterial collagen sequences (Yoshizumi et al. 2011). The two peptides of CC-Di-AB-based heterodimer have been joined with flexible linkers to make A-linker-B systems that form different assemblies, depending on the intervening linker (Boyle et al. 2012). Short linkers give fibrous materials and large colloidal assemblies, whereas longer linkers provide flexibility, allowing closure to smaller, possibly square and triangular, assemblies. Jerala and colleagues have taken this *peptide-origami* approach an elegant step further (Gradisar et al. 2013), combining eight coiled-coil pairs in a single polypeptide chain to direct the folding of a novel tetrahedron. Whilst this assembly resists high-resolution structure determination, split-protein constructs and AFM studies are fully consistent with the design.

Kros has developed a SNARE-inspired concept for peptide-directed vesicle fusion using coiled-coil peptides (Marsden et al. 2009, 2011). Through a strong and impressive series of papers, this team has shown that the early designs for "EK" heterodimer peptides from Hodges (Litowski and Hodges 2002) can be appended with lipid anchors, inserted into two different populations of lipid vesicles, which can then be mixed to effect membrane fusion in anything from simple *in vitro* systems through to live-animal models. In addition, the peptide and lipid-based components can be varied to alter and better understand the membrane-fusion events (Versluis et al. 2013a, b; Zheng et al. 2013; Schwenen et al. 2015; Kong et al. 2016; Mora et al. 2016; Zheng et al. 2016).

In the context of more-traditional synthetic biology, as mentioned above, Lim and Keating have used designer coiled coils to control transcription (Thompson et al. 2012); CC-D-AB systems with variable association constants (Thomas et al. 2013) are being used to similar effect (Smith, Savery and Woolfson, unpublished); and Swainsbury and Jones have used the CC-Di, CC-Tri and CC-Tet to bring together parts of the photosynthetic reaction centre in membranes (unpublished).

Moving onto peptide-based materials:

First, although there has been considerable progress in the development and application of coiled coil-based fibrous materials and hydrogels (Dong et al. 2008; Banwell et al. 2009; Sharp et al. 2012) this is a heavily reviewed area (Boyle and Woolfson 2011; Dasgupta et al. 2013; Woolfson and Mahmoud 2010), and will not be dwelt on further here.

Conticello has demonstrated that coiled coils can be used to make peptide nanotubes (PNTs) (Xu et al. 2013). Specifically, these assemblies use the aforementioned GCN4-based heptamer (Liu et al. 2006a), which has spiralling "stepped ends" that facilitate end-to-end assembly, in what Conticello refers to as a "lock-

washer" model. His redesign promotes assembly further, with oppositely charged groups and partly exposed hydrophobic surfaces at the termini. Inspired by this, the coiled-coil toolkit, with the exception of CC-Di—*i.e.*, the CC-Tri through CC-Hept structures—can be engineered to make peptide fibres (for CC-Tri and a variant of CC-Tet) and PNTs (CC-Pent, CC-Hex and CC-Hept) (Burgess et al. 2015). However, in these cases, the redesigned coiled coils can be blunt-ended, apparently negating the needed for ragged ends. The CC-Hex-based assemblies are particularly well ordered, allowing detailed structural studies by cryoTEM (Burgess et al. 2015). Both studies show that the PNTs can bind long-aspect ratio dyes within their lumens (Xu et al. 2013; Burgess et al. 2015), and it has been demonstrated that tubes corresponding to a single coiled-coil hexamer in diameter can be engineered (Thomas et al. 2016). Between these pieces of work, Montclare described fibres made from variants of the natural pentameric COMP, that also bind small molecules (Hume et al. 2014). These new materials hold promise for the construction of materials that could sequester and/or deliver appropriately sized small molecules.

On a related theme, Conticello and Egelman have shown that larger-diameter tubes of designed coiled coils, possibly comprising sheet-like assemblies, can be made and probed to high resolution using electron microscopy (Egelman et al. 2015). This is somewhat related to the aforementioned studies of Grigoryan and DeGrado on the decoration of carbon nanotubes with computationally designed amphipathic coiled-coil peptides (Grigoryan et al. 2011).

More complex still, a concept for Self-Assembled peptide caGEs (SAGEs) using the coiled-coil basis set has been developed as follows (Fletcher et al. 2013): first, two hubs are created (Hub-A and Hub-B). Each of these has a central CC-Tri module, with one of the outer *f* positions of each peptide functionalised as cysteine. The latter are used to couple either CC-Di-A or CC-Di-B, which have corresponding *f* sites as Cys, via disulphide bonds. In solution, these hubs behave as discrete and partly folded trimeric units. When mixed, the two hubs combine to form peptide arrays, which are assumed to be hexagonal, and then fold over and close to make spherical objects of approximately 100 nm diameter. Biophysical and microscopic analyses, and computational atomistic modelling conducted thus far indicate that the SAGEs are hollow, unilamellar, cage-like objects with defined inner and outer surfaces. They provide a potential platform for developing functional designs for the active encapsulation and delivery of drugs and biologics, and for the presentation of multiple antigenic peptides and proteins. Related to this, Marsh has constructed smaller cage-like protein objects by combining symmetry axes of coiled-coil units and natural proteins (Patterson et al. 2014). This is also reminiscent of the broader area of what might be termed directed assembly of protein-based objects, lattices and materials first proposed by Padilla and Yeates (Padilla et al. 2001), and more recently developed by Sinclair and Noble (Sinclair et al. 2011) and Yeates and Baker (King et al. 2012, 2014; Bale et al. 2015), which has been reviewed by others (Sinclair 2013; Kocar et al. 2015; Norn and Andre 2016).

2.10   Concluding Remarks

The initial aim for this chapter was to provide a comprehensive update to my 2005 review on coiled-coil design (Woolfson 2005). However, topics not covered in detail, or indeed not covered at all, include: the design and application of coiled coil-based fibrous materials (Woolfson and Mahmoud 2010; Boyle and Woolfson 2011; Dasgupta et al. 2013); the incorporation of functional metal sites into coiled-coil scaffolds (Mocny and Pecoraro 2015; Tebo and Pecoraro 2015; Slope and Peacock 2016); and the development of dynamic coiled-coil systems, such as switches and self-replicating peptides (Ambroggio and Kuhlman 2006; Pagel and Koksch 2008; Wagner and Ashkenasy 2009; Bromley and Channon 2011). These topics are covered superbly well by others in the reviews cited in the preceding sentences. What this chapter does highlight, however, is that the field of coiled-coil design is brimming with scope and ideas, and that there is the talent to realise the many ambitions that come with these. If the advances that have been made over the past decade or so are anything to judge by, the future of field is looking very rosy indeed.

On this, some speculation is justified. It is anticipated that rational designs for much more-complex coiled-coil architectures will occur in the near future—that is, there will be designed assemblies with more helices, more layers, different orientations of helices, bigger barrels, and multi-component materials. It is also envisaged that useful functions will be incorporated, including specified binding sites and enzyme-like activities, into these *de novo* coiled-coil folds (Burton and Woolfson, 2016); and more-dynamic designs, with coiled coils that can switch state on cue in response to particular stimulus, are foreseen (Lizatovic et al. 2016). The hope is that all of this is done in a collaborative and increasingly computational-based environment, where software is developed to meet challenges in both basic and applied science. For instance, an ability to not only predict the most likely structure for a designed coiled-coil sequence, but also its free energy with more confidence—or at least a free energy difference from the nearest off-target design—would be extremely useful; as would computational coiled-coil and, more generally, protein-design tools that are accessible to non-experts, for example to synthetic biologists, who will have different ambitions than structural biologists.

References

Ambroggio XI, Kuhlman B (2006) Design of protein conformational switches. Curr Opin Struct Biol 16(4):525–530

Apgar JR, Gutwin KN, Keating AE (2008) Predicting helix orientation for coiled-coil dimers. Proteins 72(3):1048–1065

Armstrong CT, Vincent TL, Green PJ, Woolfson DN (2011) SCORER 2.0: an algorithm for distinguishing parallel dimeric and trimeric coiled-coil sequences. Bioinformatics 27(14):1908–1914

Aronsson C, Danmark S, Zhou F, Oberg P, Enander K, Su H, Aili D (2015) Self-sorting heterodimeric coiled coil peptides with defined and tuneable self-assembly properties. Sci Rep:5

Baker EG, Bartlett GJ, Crump MP, Sessions RB, Linden N, Faul CFJ, Woolfson DN (2015) Local and macroscopic electrostatic interactions in single alpha-helices. Nat Chem Biol 11(3):221–228

Bale JB, Park RU, Liu YX, Gonen S, Gonen T, Cascio D, King NP, Yeates TO, Baker D (2015) Structure of a designed tetrahedral protein assembly variant engineered to have improved soluble expression. Protein Sci 24(10):1695–1701

Banwell EF, Abelardo ES, Adams DJ, Birchall MA, Corrigan A, Donald AM, Kirkland M, Serpell LC, Butler MF, Woolfson DN (2009) Rational design and application of responsive alpha-helical peptide hydrogels. Nat Mater 8(7):596–600

Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, Sankaran B, Seelig G, Zwart PH, Baker D (2016) De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science 352(6286):680–687

Boyle AL, Woolfson DN (2011) De novo designed peptides for biological applications. Chem Soc Rev 40(8):4295–4306

Boyle AL, Bromley EHC, Bartlett GJ, Sessions RB, Sharp TH, Williams CL, Curmi PMG, Forde NR, Linke H, Woolfson DN (2012) Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. J Am Chem Soc 134(37):15457–15467

Bromley EHC, Channon KJ (2011) Alpha-helical peptide assemblies: giving new function to designed structures. Prog Mol Biol Transl Sci 103:231–275

Bromley EHC, Channon K, Moutevelis E, Woolfson DN (2008) Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems. ACS Chem Biol 3(1):38–50

Bromley EHC, Sessions RB, Thomson AR, Woolfson DN (2009) Designed alpha-helical tectons for constructing multicomponent synthetic biological systems. J Am Chem Soc 131(3):928–930

Brown JH, Cohen C, Parry DAD (1996) Heptad breaks in alpha-helical coiled coils: stutters and stammers. Proteins 26(2):134–145

Burgess NC, Sharp TH, Thomas F, Wood CW, Thomson AR, Zaccai NR, Brady RL, Serpell LC, Woolfson DN (2015) Modular design of self-assembling peptide-based nanotubes. J Am Chem Soc 137(33):10554–10562

Burton AJ, Thomas F, Agnew C, Hudson KL, Halford SE, Brady RL, Woolfson DN (2013) Accessibility, reactivity, and selectivity of side chains within a channel of de novo peptide assembly. J Am Chem Soc 135(34):12524–12527

Burton AJ, Thomson AR, Dawson WM, Brady RL, Woolfson DN (2016) Installing hydrolytic activity into a completely de novo protein framework. Nat Chem 8(9):837–844.

Channon K, Bromley EHC, Woolfson DN (2008) Synthetic biology through biomolecular design and engineering. Curr Opin Struct Biol 18(4):491–498

Chothia C, Levitt M, Richardson D (1981) Helix to helix packing in proteins. J Mol Biol 145(1):215–250

Crick FHC (1953a) The Fourier transform of a coiled-coil. Acta Crystallogr 6(8-9):685–689

Crick FHC (1953b) The packing of alpha-helices – simple coiled-coils. Acta Crystallogr 6(8-9):689–697

Crooks RO, Baxter D, Panek AS, Lubben AT, Mason JM (2016) Deriving heterospecific self-assembling protein-protein interactions using a computational interactome screen. J Mol Biol 428(2):385–398

Dasgupta A, Mondal JH, Das D (2013) Peptide hydrogels. RSC Adv 3(24):9117–9149

Davey JA, Chica RA (2012) Multistate approaches in computational protein design. Protein Sci 21(9):1241–1252

Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18(4):617–625

Diss ML, Kennan AJ (2008) Heterotrimeric coiled coils with core residue urea side chains. J Org Chem 73(24):9752–9755

Dong CJ, Beis K, Nesper J, Brunkan-LaMontagne AL, Clarke BR, Whitfield C, Naismith JH (2006) Wza the translocon for E-coli capsular polysaccharides defines a new class of membrane protein. Nature 444(7116):226–229

Dong H, Paramonov SE, Hartgerink JD (2008) Self-assembly of alpha-helical coiled coil nanofibers. J Am Chem Soc 130(41):13691–13695

Egelman EH, Xu C, DiMaio F, Magnotti E, Modlin C, Yu X, Wright E, Baker D, Conticello VP (2015) Structural plasticity of helical nanotubes based on coiled-coil assemblies. Structure 23(2):280–289

Eshaghi S, Niegowski D, Kohl A, Molina DM, Lesley SA, Nordlund P (2006) Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution. Science 313(5785):354–357

Fairman R, Chao HG, Lavoie TB, Villafranca JJ, Matsueda GR, Novotny J (1996) Design of heterotetrameric coiled coils: evidence for increased stabilization by Glu(-)-Lys(+) ion pair interactions. Biochemistry 35(9):2824–2829

Feldmeier K, Hocker B (2013) Computational protein design of ligand binding and catalysis. Curr Opin Chem Biol 17(6):929–933

Fletcher JM, Boyle AL, Bruning M, Bartlett GJ, Vincent TL, Zaccai NR, Armstrong CT, Bromley EHC, Booth PJ, Brady RL, Thomson AR, Woolfson DN (2012) A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. ACS Synth Biol 1(6):240–250

Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, Miles MJ, Sessions RB, Verkade P, Woolfson DN (2013) Self-assembling cages from coiled-coil peptide modules. Science 340(6132):595–599

Fong JH, Keating AE, Singh M (2004) Predicting specificity in bZIP coiled-coil protein interactions. Genome Biol 5(2): R11

Ghirlanda G, Lear JD, Ogihara NL, Eisenberg D, DeGrado WF (2002) A hierarchic approach to the design of hexameric helical barrels. J Mol Biol 319(1):243–253

Ghosh I, Hamilton AD, Regan L (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. J Am Chem Soc 122(23):5658–5659

Gonzalez L, Woolfson DN, Alber T (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat Struct Biol 3(12):1011–1018

Gradisar H, Jerala R (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. J Pept Sci 17(2):100–106

Gradisar H, Bozic S, Doles T, Vengust D, Hafner-Bratkovic I, Mertelj A, Webb B, Sali A, Klavzar S, Jerala R (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nat Chem Biol 9(6):362–366

Grigoryan G, DeGrado WF (2011) Probing designability via a generalized model of helical bundle geometry. J Mol Biol 405(4):1079–1100

Grigoryan G, Kim YH, Acharya R, Axelrod K, Jain RM, Willis L, Drndic M, Kikkawa JM, DeGrado WF (2011) Computational design of virus-like protein assemblies on carbon nanotube surfaces. Science 332(6033):1071–1076

Gruber M, Lupas AN (2003) Historical review: another 50th anniversary – new periodicities in coiled coils. Trends Biochem Sci 28(12):679–685

Gurnon DG, Whitaker JA, Oakley MG (2003) Design and characterization of a homodimeric anti-parallel coiled coil. J Am Chem Soc 125(25):7518–7519

Hadley EB, Gellman SH (2006) An antiparallel alpha-helical coiled-coil model system for rapid assessment of side-chain recognition at the hydrophobic interface. J Am Chem Soc 128(51):16444–16445

Hadley EB, Testa OD, Woolfson DN, Gellman SH (2008) Preferred side-chain constellations at antiparallel coiled-coil interfaces. Proc Natl Acad Sci U S A 105(2):530–535

Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between 2-stranded, 3-stranded and 4-stranded coiled coils in Gcn4 leucine-zipper mutants. Science 262(5138):1401–1407

Harbury PB, Kim PS, Alber T (1994) Crystal-structure of an isoleucine-zipper trimer. Nature 371(6492):80–83

Harbury PB, Tidor B, Kim PS (1995) Repacking protein cores with backbone freedom – structure prediction for coiled coils. Proc Natl Acad Sci U S A 92(18):8408–8412

Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS (1998) High-resolution protein design with back-bone freedom. Science 282(5393):1462–1467

Hartmann MD, Ridderbusch O, Zeth K, Albrecht R, Testa O, Woolfson DN, Sauer G, Dunin-Horkawicz S, Lupas AN, Alvarez BH (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. Proc Natl Acad Sci U S A 106(40):16950–16955

Hicks MR, Holberton DV, Kowalczyk C, Woolfson DN (1997) Coiled-coil assembly by peptides with non-heptad sequence motifs. Fold Des 2(3):149–158

Hicks MR, Walshaw J, Woolfson DN (2002) Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. J Struct Biol 137(1-2):73–81

Hol WGJ, Vanduijnen PT, Berendsen HJC (1978) Alpha-helix dipole and properties of proteins. Nature 273(5662):443–446

Huang PS, Oberdorfer G, Xu CF, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B, Baker D (2014) High thermodynamic stability of parametrically designed helical bundles. Science 346(6208):481–485

Hume J, Sun J, Jacquet R, Renfrew PD, Martin JA, Bonneau R, Gilchrist ML, Montclare JK (2014) Engineered coiled-coil protein microfibers. Biomacromolecules 15(10):3503–3510

Joh NH, Wang T, Bhate MP, Acharya R, Wu YB, Grabe M, Hong M, Grigoryan G, DeGrado WF (2014) De novo design of a transmembrane Zn2+-transporting four-helix bundle. Science 346(6216):1520-1524.

Kaplan JB, Reinke AW, Keating AE (2014) Increasing the affinity of selective bZIP-binding pep-tides through surface residue redesign. Protein Sci 23(7):940–953

Kashiwada A, Hiroaki H, Kohda D, Nango M, Tanaka T (2000) Design of a heterotrimeric alpha-helical bundle by hydrophobic core engineering. J Am Chem Soc 122(2):212–215

King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, Gonen T, Yeates TO, Baker D (2012) Computational design of self-assembling protein nanomaterials with atomic level accu-racy. Science 336(6085):1171–1174

King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D (2014) Accurate design of co-assembling multi-component protein nanomaterials. Nature 510(7503):103–108

Kiyokawa T, Kanaori K, Tajima K, Kawaguchi M, Mizuno T, Oku JI, Tanaka T (2004) Selective formation of AAB- and ABC-type heterotrimeric alpha-helical coiled coils. Chem Eur J 10(14):3548–3554

Kocar V, Abram SB, Doles T, Basic N, Gradisar H, Pisanski T, Jerala R (2015) TOPOFOLD, the designed modular biomolecular folds: polypeptide-based molecular origami nanostructures following the footsteps of DNA. Wiley Interdiscip Rev Nanomed Nanobiotechnol 7(2):218–237

Kong L, Askes SHC, Bonnet S, Kros A, Campbell F (2016) Temporal control of membrane fusion through photolabile PEGylation of liposome membranes. Angew Chem Int Ed 55(4):1396–1400

Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature 405(6789):914–919

Litowski JR, Hodges RS (2002) Designing heterodimeric two-stranded alpha-helical coiled-coils – effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. J Biol Chem 277(40):37272–37279

Liu J, Zheng Q, Deng Y, Cheng C-S, Kallenbach NR, Lu M (2006a) A seven-helix coiled coil. Proc Natl Acad Sci U S A 103(42):15457–15462

Liu J, Zheng Q, Deng YQ, Kallenbach NR, Lu M (2006b) Conformational transition between four and five-stranded phenylalanine zippers determined by a local packing interaction. J Mol Biol 361(1):168–179

Lizatovic R, Aurelius O, Stenstrom O, Drakenberg T, Akke M, Logan DT, Andre I (2016) A de novo designed coiled-coil peptide with a reversible pH-induced oligomerization switch.

Lovejoy B, Choe S, Cascio D, McRorie DK, Degrado WF, Eisenberg D (1993) Crystal-structure of a synthetic triple-stranded alpha-helical bundle. Science 259(5099):1288–1293

Lumb KJ, Kim PS (1998) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil (vol 34, pg 8642, 1995). Biochemistry 37(37):13042–13042

Lupas A (1996a) Coiled coils: new structures and new functions. Trends Biochem Sci 21(10):375–382

Lupas A (1996b) Prediction and analysis of coiled-coil structures. Methods Enzymol 266:513–525

Lupas AN, Gruber M (2005) The structure of alpha-helical coiled coils. Adv Protein Chem 70:37–78

Lupas A, Vandyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252(5009):1162–1164

Magliery TJ (2015) Protein stability: computation, sequence statistics, and new experimental methods. Curr Opin Struct Biol 33:161–168

Malashkevich VN, Kammerer RA, Efimov VP, Schulthess T, Engel J (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel? Science 274(5288):761–765

Marsden HR, Elbers NA, Bomans PHH, Sommerdijk NAJM, Kros A (2009) A reduced SNARE model for membrane fusion. Angew Chem Int Ed 48(13):2330–2333

Marsden HR, Tomatsu I, Kros A (2011) Model systems for membrane fusion. Chem Soc Rev 40(3):1572–1585

McClain DL, Woods HL, Oakley MG (2001) Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. J Am Chem Soc 123(13):3151–3152

McDonnell AV, Jiang T, Keating AE, Berger B (2006) Paircoil2: improved prediction of coiled coils from sequence. Bioinformatics 22(3):356–358

Mizuno T, Suzuki K, Imai T, Kitade Y, Furutani Y, Kudou M, Oda M, Kandori H, Tsumoto K, Tanaka T (2009) Manipulation of protein-complex function by using an engineered heterotrimeric coiled-coil switch. Org Biomol Chem 7(15):3102–3111

Mocny CS, Pecoraro VL (2015) De novo protein design as a methodology for synthetic bioinorganic chemistry. Acc Chem Res 48(8):2388–2396

Moll JR, Ruvinov SB, Pastan I, Vinson C (2001) Designed heterodimerizing leucine zippers with a range of pIs and stabilities up to 10(-15) M. Protein Sci 10(3):649–655

Monera OD, Zhou NE, Kay CM, Hodges RS (1993) Comparison of antiparallel and parallel 2-stranded alpha-helical coiled-coils – design, synthesis, and characterization. J Biol Chem 268(26):19218–19227

Monera OD, Kay CM, Hodges RS (1994) Electrostatic interactions control the parallel and anti-parallel orientation of alpha-helical chains in 2-stranded alpha-helical coiled-coils. Biochemistry 33(13):3862–3871

Mora NL, Bahreman A, Valkenier H, Li HY, Sharp TH, Sheppard DN, Davis AP, Kros A (2016) Targeted anion transporter delivery by coiled-coil driven membrane fusion. Chem Sci 7(3):1768–1772

Moutevelis E, Woolfson DN (2009) A periodic table of coiled-coil protein structures. J Mol Biol 385(3):726–732

Myszka DG, Chaiken IM (1994) Design and characterization of an intramolecular antiparallel coiled-coil peptide. Biochemistry 33(9):2363–2372

Nautiyal S, Alber T (1999) Crystal structure of a designed, thermostable; heterotrimeric coiled coil. Protein Sci 8(1):84–90

Nautiyal S, Woolfson DN, King DS, Alber T (1995) A designed heterotrimeric coiled-coil. Biochemistry 34(37):11645–11651

Negron C, Keating AE (2013) Multistate protein design using CLEVER and CLASSY. Methods Enzymol 523:171–190

Negron C, Keating AE (2014) A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit. J Am Chem Soc 136(47):16544–16556

Norn CH, Andre I (2016) Computational design of protein self-assembly. Curr Opin Struct Biol 39:39–45

Oakley MG, Hollenbeck JJ (2001) The design of antiparallel coiled coils. Curr Opin Struct Biol 11(4):450–457

Offer G, Sessions R (1995) Computer modeling of the alpha-helical coiled-coil – packing of side-chains in the inner-core. J Mol Biol 249(5):967–987

Offer G, Hicks MR, Woolfson DN (2002) Generalized Crick equations for modeling noncanonical coiled coils. J Struct Biol 137(1-2):41–53

Ogihara NL, Weiss MS, Degrado WF, Eisenberg D (1997) The crystal structure of the designed trimeric coiled coil-V(a)L(d): implications for engineering crystals and supramolecular assemblies. Protein Sci 6(1):80–88

Oshaben KM, Salari R, McCaslin DR, Chong LT, Horne WS (2012) The native GCN4 leucine-zipper domain does not uniquely specify a dimeric oligomerization state. Biochemistry 51(47):9581–9591

O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the Gcn4 leucine zipper, a 2-stranded, parallel coiled coil. Science 254(5031):539–544

O'Shea EK, Lumb KJ, Kim PS (1993) Peptide velcro – design of a heterodimeric coiled-coil. Curr Biol 3(10):658–667

Padilla JE, Colovos C, Yeates TO (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc Natl Acad Sci U S A 98(5):2217–2221

Pagel K, Koksch B (2008) Following polypeptide folding and assembly with conformational switches. Curr Opin Chem Biol 12(6):730–739

Pandya MJ, Cerasoli E, Joseph A, Stoneman RG, Waite E, Woolfson DN (2004) Sequence and structural duality: designing peptides to adopt two stable conformations. J Am Chem Soc 126(51):17016–17024

Patterson DP, Su M, Franzmann TM, Sciore A, Skiniotis G, Marsh ENG (2014) Characterization of a highly flexible self-assembling protein system designed to form nanocages. Protein Sci 23(2):190–199

Payandeh J, Pai EF (2006) A structural basis for Mg2+ homeostasis and the CorA translocation cycle. EMBO J 25(16):3762–3773

Plecs JJ, Harbury PB, Kim PS, Alber T (2004) Structural test of the parameterized-backbone method for protein design. J Mol Biol 342(1):289–297

Potapov V, Kaplan JB, Keating AE (2015) Data-driven prediction and design of bZIP coiled-coil interactions. PLoS Comput Biol 11(2) UNSP e1004046

Quinlan RA, Bromley EH, Pohl E (2015) A silk purse from a sow's ear – bioinspired materials based on alpha-helical coiled coils. Curr Opin Cell Biol 32:131–137

Rackham OJL, Madera M, Armstrong CT, Vincent TL, Woolfson DN, Gough J (2010) The evolution and structure prediction of coiled coils across all genomes. J Mol Biol 403(3):480–493

Ramisch S, Lizatovic R, Andre I (2015) Exploring alternate states and oligomerization preferences of coiled-coils by de novo structure modeling. Proteins 83(2):235–247

Regan L, Caballero D, Hinrichsen MR, Virrueta A, Williams DM, O'Hern CS (2015) Protein design: past, present, and future. Biopolymers 104(4):334–350

Reinke AW, Grant RA, Keating AE (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. J Am Chem Soc 132(17):6025–6031

Root BC, Pellegrino LD, Crawford ED, Kokona B, Fairman R (2009) Design of a heterotetrameric coiled coil. Protein Sci 18(2):329–336

Rose PW, Prlic A, Bi CX, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woo J, Young J, Zardecki C, Berman HM, Bourne PE, Burley SK (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 43(D1):D345–D356

Sastri NP, Viskovska M, Hyser JM, Tanner MR, Horton LB, Sankaran B, Prasad BVV, Estes MK (2014) Structural plasticity of the coiled-coil domain of rotavirus NSP4. J Virol 88(23):13602–13612

Schnarr NA, Kennan AJ (2002) Peptide tic-tac-toe: heterotrimeric coiled-coil specificity from steric matching of multiple hydrophobic side chains. J Am Chem Soc 124(33):9779–9783

Schwenen LLG, Hubrich R, Milovanovic D, Geil B, Yang J, Kros A, Jahn R, Steinem C (2015) Resolving single membrane fusion events on planar pore-spanning membranes. Sci Rep 5

Sharp TH, Bruning M, Mantell J, Sessions RB, Thomson AR, Zaccai NR, Brady RL, Verkade P, Woolfson DN (2012) Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of de novo design. Proc Natl Acad Sci U S A 109(33):13266–13271

Sinclair JC (2013) Constructing arrays of proteins. Curr Opin Chem Biol 17(6):946–951

Sinclair JC, Davies KM, Venien-Bryan C, Noble MEM (2011) Generation of protein lattices by fusing proteins with matching rotational symmetry. Nat Nanotechnol 6(9):558–562

Slope LN, Peacock AFA (2016) De novo design of xeno-metallo coiled coils. Chem Asian J 11(5):660–666

Steinkruger JD, Bartlett GJ, Hadley EB, Fay L, Woolfson DN, Gellman SH (2012a) The d′-d-d′ vertical triad is less discriminating than the a′-a-a′ vertical triad in the antiparallel coiled-coil dimer motif. J Am Chem Soc 134(5):2626–2633

Steinkruger JD, Bartlett GJ, Woolfson DN, Gellman SH (2012b) Strong contributions from vertical triads to helix-partner preferences in parallel coiled coils. J Am Chem Soc 134(38):15652–15655

Strelkov SV, Burkhard P (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J Struct Biol 137(1-2):54–64

Sun L, Young LN, Zhang XZ, Boudko SP, Fokine A, Zbornik E, Roznowski AP, Molineux IJ, Rossmann MG, Fane BA (2014) Icosahedral bacteriophage Phi X174 forms a tail for DNA transport during infection. Nature 505(7483):432–435

Surkont J, Diekmann Y, Ryder PV, Pereira-Leal JB (2015) Coiled-coil length: size does matter. Proteins 83(12):2162–2169

Tanaka Y, Sasaki T, Kumagai I, Yasutake Y, Yao M, Tanaka I, Tsumoto K (2007) Molecular properties of two proteins homologous to PduO-Type ATP: cob(I)alamin adenosyltransferase from Sulfolobus tokodaii. Proteins 68(2):446–457

Taylor WR, Chelliah V, Hollup SM, MacDonald JT, Jonassen I (2009) Probing the "Dark Matter" of protein fold space. Structure 17(9):1244–1252

Tebo AG, Pecoraro VL (2015) Artificial metalloenzymes derived from three-helix bundles. Curr Opin Chem Biol 25:65–70

Testa OD, Moutevelis E, Woolfson DN (2009) CC plus: a relational database of coiled-coil structures. Nucleic Acids Res 37:D315-D322.

Thomas F, Boyle AL, Burton AJ, Woolfson DN (2013) A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime. J Am Chem Soc 135(13):5161–5166

Thomas F, Burgess NC, Thomson AR, Woolfson DN (2016) Controlling the assembly of coiled-coil peptide nanotubes. Angew Chem Int Ed 55(3):987–991

Thompson KE, Bashor CJ, Lim WA, Keating AE (2012) SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. ACS Synth Biol 1(4):118–129

Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble alpha-helical barrels. Science 346(6208):485–488

Trigg J, Gutwin K, Keating AE, Berger B (2011) Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. PLoS One 6(8) e23519

Versluis F, Dominguez J, Voskuhl J, Kros A (2013a) Coiled-coil driven membrane fusion: zipper-like vs. non-zipper-like peptide orientation. Faraday Discuss 166:349–359

Versluis F, Voskuhl J, van Kolck B, Zope H, Bremmer M, Albregtse T, Kros A (2013b) In situ modification of plain liposomes with lipidated coiled coil forming peptides induces membrane fusion. J Am Chem Soc 135(21):8057–8062

Vincent TL, Green PJ, Woolfson DN (2013) LOGICOIL-multi-state prediction of coiled-coil oligomeric state. Bioinformatics 29(1):69–76

Vostrikov VV, Mote KR, Verardi R, Veglia G (2013) Structural dynamics and topology of phosphorylated phospholamban homopentamer reveal its role in the regulation of calcium transport. Structure 21(12):2119–2130

Wagner N, Ashkenasy G (2009) Systems chemistry: logic gates, arithmetic units, and network motifs in small networks. Chem-Eur J 15(7):1765–1775

Walshaw J, Woolfson DN (2001a) Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. Protein Sci 10(3):668–673

Walshaw J, Woolfson DN (2001b) SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures. J Mol Biol 307(5):1427–1450

Walshaw J, Woolfson DN (2003) Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. J Struct Biol 144(3):349–361

Walther D, Eisenhaber F, Argos P (1996) Principles of helix-helix packing in proteins: the helical lattice superposition model. J Mol Biol 255(3):536–553

Wood CW, Bruning M, Ibarra AA, Bartlett GJ, Thomson AR, Sessions RB, Brady RL, Woolfson DN (2014) CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinformatics 30(21):3029–3035

Woolfson DN (2005) The design of coiled-coil structures and assemblies. Adv Protein Chem 70:79–112

Woolfson DN, Alber T (1995) Predicting oligomerization states of coiled coils. Protein Sci 4(8):1596–1607

Woolfson DN, Mahmoud ZN (2010) More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. Chem Soc Rev 39(9):3464–3479

Woolfson DN, Bartlett GJ, Bruning M, Thomson AR (2012) New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. Curr Opin Struct Biol 22(4):432–441

Woolfson DN, Bartlett GJ, Burton AJ, Heal JW, Niitsu A, Thomson AR, Wood CW (2015) De novo protein design: how do we expand into the universe of possible protein structures? Curr Opin Struct Biol 33:16–26

Xu C, Liu R, Mehta AK, Guerrero-Ferreira RC, Wright ER, Dunin-Horkawicz S, Morris K, Serpell LC, Zuo X, Wall JS, Conticello VP (2013) Rational design of helical nanotubes from self-aof coiled-coil lock washers. J Am Chem Soc 135(41):15565–15578

Yoshizumi A, Fletcher JM, Yu ZX, Persikov AV, Bartlett GJ, Boyle AL, Vincent TL, Woolfson DN, Brodsky B (2011) Designed coiled coils promote folding of a recombinant bacterial collagen. J Biol Chem 286(20):17512–17520

Zaccai NR, Chi B, Thomson AR, Boyle AL, Bartlett GJ, Bruning M, Linden N, Sessions RB, Booth PJ, Brady RL, Woolfson DN (2011) A de novo peptide hexamer with a mutable channel. Nat Chem Biol 7(12):935–941

Zheng TT, Voskuhl J, Versluis F, Zope HR, Tomatsu I, Marsden HR, Kros A (2013) Controlling the rate of coiled coil driven membrane fusion. Chem Commun 49(35):3649–3651

Zheng TT, Bulacu M, Daudey G, Versluis F, Voskuhl J, Martelli G, Raap J, Sevink GJA, Kros A, Boyle AL (2016) A non-zipper-like tetrameric coiled coil promotes membrane fusion. RSC Adv 6(10):7990–7998

Chapter 3
Functional and Structural Roles of Coiled Coils

Marcus D. Hartmann

Contents

Abstract Coiled coils appear in countless structural contexts, as appendages to small proteins, as parts of multi-domain proteins, and as building blocks of filaments. Although their structure is unpretentious and their basic properties are understood in great detail, the spectrum of functional properties they provide in different proteins has become increasingly complex. This chapter aims to depict this functional spectrum, to identify common themes and their molecular basis, with an emphasis on new insights gained into dynamic aspects.

M.D. Hartmann (✉)
Max Planck Institute for Developmental Biology,
Spemannstraße 35, 72076 Tübingen, Germany
e-mail: Marcus.Hartmann@tuebingen.mpg.de

Keywords α-helix • Oligomerization • Molecular recognition • Molecular spacer • Trimeric autotransporter adhesin • Surface proteins • Polar core residues • Coiled-coil periodicity • *β*-layer • Protein folding and disaggregation • Unfolding and degradation • Two-component signal transduction • Histidine kinases • Methyl-accepting chemoreceptors • Asymmetry

Abbreviations

AAA	ATPases associated with diverse cellular activities
ARC	ATPase forming Ring-shaped Complexes
b/ZIP	basic region leucine zipper
CA domain	catalytic domain
CCT	Chaperonin-Containing TCP1
ClpB	Caseinolytic protease component B
DHp domain	Dimerization and Histidine phosphotransfer domain
HAMP domain	domain occurring in Histidine kinases, Adenylyl cyclases, Methyl-accepting chemotaxis proteins, and Phosphatases
HslU	heat shock locus U
HslV	heat shock locus V
Hsp70	70 kilodalton heat shock protein
IF	intermediate filament
NadA	Neisserial adhesin A
p-helix	periplasmic helix
PAN	Proteasome Activating Nucleotidase
PFD	prefoldin
RPT	Regulatory Particle Triple-A ATPase
Skp	Seventeen-kilodalton protein
SMC	structural maintenance of chromosomes
SNARE	soluble N-ethyl maleimide sensitive factor attachment protein receptor
TAA	trimeric autotransporter adhesin

3.1 The Basic Architecture of Coiled Coils

The α-helical coiled coil is an omnipresent protein fold, accounting for about 3 per cent of all protein-coding regions across all genomes (Rackham et al. 2010) and contributing to an ever-growing list of functional contexts (Lupas 1996; Burkhard et al. 2001; Parry 2014). A major hallmark of coiled coils is their structural simplicity: they consist of two or more α-helices that are wound around each other into supercoils with a common hydrophobic core, while the packing of the helices and the side-chains in the core satisfies simple rules. In their simplest form, they are parallel homo-oligomers with a seven-residue sequence repeat pattern, the heptad repeat. Therein, the seven individual positions are labelled *a*, *b*, *c*, *d*, *e*, *f*, *g*, of which

positions *a* an *d* contribute to the core and are thus mostly hydrophobic (Fig. 3.1). The seven residues of a single repeat are accommodated in two helical turns with respect to the coiled-coil axis; as the resulting 3.5 residues per turn are less than the 3.62 residues per turn of unperturbed α-helices, the whole bundle has a left-handed twist. Interestingly, in addition to the heptad repeat, several related sequence periodicities also exist. They can be explained by short insertions into the heptad repeat that lead to different core packings and can result in rather straight or right-handed supercoils (Lupas and Gruber 2005; Hartmann et al. 2016). Widespread examples are the eleven-residue hendecad and the 15-residue pentadecad repeats. A hendecad (a.k.a. undecad) can be brought about by the insertion of four residues into the heptad repeat – such an insertion of four residues is called a stutter. Its eleven residues are accommodated in three helical turns, yielding a slightly right-handed bundle with 3.67 residues per turn. Likewise, a pentadecad can be seen as resulting from the insertion of two stutters. With 15 residues accommodated in four helical turns, it has a periodicity of 3.75 and thus a more pronounced right-handedness (Brown et al. 1996; Strelkov and Burkhard 2002; Parry et al. 2008). As the four residues of stutters are more than the 3.62 residues of one helical turn, they generally increase right-handedness. The opposite effect is achieved by the insertion of three residue elements – so called stammers. In essence, all periodicities accessible to coiled coils can be described by the insertion of one or more stammers or stutters into the heptad repeat, adjusting the winding mode of the coiled coil locally or globally.

The individual positions in the repeat sequence have different amino-acid preferences and play different roles in coiled-coil architecture. The stability and oligomerization specificity of coiled coils, for instance, is largely dependent on the nature of the core residues. In heptads, the core packing is determined by positions *a* and *d* that assume a knobs-into-holes pattern, in which a core residue from one helix is accommodated in a "hole" formed by four residues of the other helix, resulting in a side-by-side arrangement of the hydrophobic core residues (Lupas and Gruber 2005). This packing mode yields different geometrical preferences for core residues in different oligomeric states. This was impressively demonstrated in a comparative study using the GCN4 leucine zipper as a model, in which different sets of core
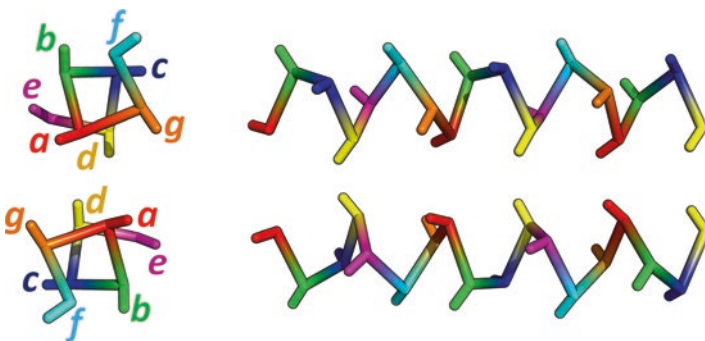
Fig. 3.1 The seven positions in a dimeric heptad coiled coil

residues could switch the oligomerization state between dimeric (isoleucines in *a*, leucines in *d*), tetrameric (leucines in *a*, isoleucines in *d*) and trimeric (other combinations) (Harbury et al. 1993). However, in non-heptad periodicities, the core packing is more complex and involves additional packing modes different from knobs-into-holes. These involve the so-called da- and x-layers, in which the core residues form either rings of interactions around the bundle axis or all point their side-chains towards the bundle axis, respectively, that have yet different amino-acid preferences (Lupas and Gruber 2005). Apart from hydrophobic core residues, polar and charged residues play important roles in coiled-coil architecture. In core positions, polar residues can influence oligomerization specificity – usually at the cost of stability (Eckert et al. 1998; Hartmann et al. 2009). In other positions, they can have the opposite effect and increase stability. For example, positions *e* and *g*, which are adjacent to the hydrophobic core and often rich in charged residues, frequently form stabilizing inter-helical salt bridges. Thus, especially in hetero-oligomeric coiled coils, positions *e* and *g* play an important role in pairing specificity.

Generally, owing to their simple architecture, the theoretical concept of coiled coils is so well understood that their structure can usually be predicted and computed from sequence – and often also designed - with great confidence (Lupas et al. 1991; Lupas and Gruber 2005; Woolfson 2005; Wood et al. 2014). Comprehensive descriptions of the current knowledge of coiled-coil structure and the state-of-the-art in coiled-coil design are found in Chaps. 4 (Lupas et al. 2017) and 2 (Woolfson 2017), respectively.

For the present chapter, we want to discuss two key observations:

(1) The supercoil parameters governing the way the helices are wound in the bundle are set by the sequence repeat length, and can be altered by insertions into the heptad repeat. (2) The stability, the oligomerization specificity, and the folding propensity is tuned by the nature of amino acids in the individual positions of the repeat sequence, especially those inside and peripheral to the core.

3.2 Classical Roles of Coiled Coils

3.2.1 Filament-Forming Coiled-Coil Proteins

The field of coiled-coil research was initiated through X-ray diffraction experiments performed throughout the 1930s and 1940s on keratin-, myofilament-, and related protein fibres, that all revealed similar features in their diffraction patterns. Later, in the 1950s, Pauling constructed a model of what is now known as the α-helix, which could explain most features of these diffraction patterns. However, it was soon understood by both Pauling and Crick that these helices needed to be assembled into a superhelix with a certain tilting angle if the patterns were to be fully explained. This provided the basis of the coiled-coil standard model (reviewed in Lupas and Gruber 2005; Parry et al. 2008). As protein fibres and fibrous proteins dominated by a single and simple structural motif, like the heptad repeat of coiled coils, posed ideal samples for diffraction experiments, this class of proteins dominated the first decades of coiled-coil research.

Tropomyosin (1C1G)

Kinesin (1N6M)   **FOS•JUN•DNA (1FOS)**   **SNARE complex (3IPD)**
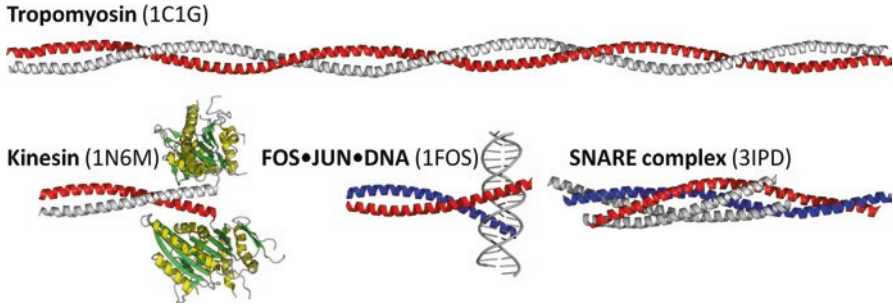
Fig. 3.2 Classical coiled-coil proteins. For the homodimeric Tropomyosin and Kinesin, one helix of the coiled coil is shown in *red*. The two chains of the heterodimeric Jun:Fos complex and the three chains of the heterotetrameric SNARE complex are coloured individually

The first characterized coiled-coil protein was the intermediate filament (IF) protein α-keratin. In contrast to the thinner microfilaments and the thicker microtubules that are composed of only one (actin) or two (α- and β-tubulin) globular building blocks, the architecture of the IF is rather complex. The basic building blocks of the IF are dimeric proteins, which are grouped into six main types. Depending on the type, these dimers are either homo- or heterotypic, but all of them follow the same architecture: they start with an N-terminal "head" domain, pass into an extended parallel coiled-coil "rod" domain that mediates the dimerization, and end in the C-terminal "tail" domain (Parry 2005). Whereas the head and tail domains appear to be intrinsically disordered in most IF proteins *in vitro* (Guharoy et al. 2013), the rod domain follows essentially the same organization in all IF protein types. It is an almost continuous, mostly heptad coiled coil that is – depending on the type – interrupted by one or two conserved, short linkers that are unlikely to be completely α-helical (Chernyatina et al. 2015). In the next level of assembly, the IF dimers are thought to form antiparallel, half-staggered tetramers. By lateral association, these tetramers form building blocks that are assembled into the mature filaments by longitudinal annealing (Herrmann et al. 2007). These filaments have astonishing mechanical properties, as they can be stretched to a multiple of their initial length (Kreplak et al. 2005). This property can be ascribed mostly to the coiled coils of the rod domains, which are able to undergo a reversible phase transition from α-helix to β-sheet (Miserez and Guerette 2013). For further details on IF structure and function, the reader is referred to Chaps. 5 (Fraser and Parry 2017), 6 (Guzenko et al. 2017) and 7 (Bouameur and Magin 2017). For the purpose of this chapter, we note that the coiled coils of the IF constitute most of its structural backbone, serving as molecular spacers of a defined length that are, however, seemingly indestructible by axial stretching because of their ability to undergo a reversible and elastic phase transition – an invaluable property for an element of the cytoskeleton.

In contrast to the IF, the core structure of the actin-containing microfilaments does not contain coiled coils itself, but coiled-coil co-proteins are crucial to virtually all its functions. A key player is tropomyosin, an integral regulatory component in both the actin-containing microfilaments of the cytoskeleton and the thin myofilaments of

muscle fibres (Fig. 3.2). A multitude of tropomyosin isoforms, resulting from different genes and splicing variants with different functions, can homo- or heterodimerize into parallel coiled coils that associate alongside the actin filaments to regulate the access and interactions of actin-binding proteins (reviewed in Gunning et al. (2015) and subject of Chap. 9 (Hitchcock-DeGregori and Barua 2017). One such class of actin-binding proteins are the dimeric myosin motor proteins, key mediators of actin-based motility and building blocks of the thick myofilaments. Each half of the myosin dimer consists of a "head" domain that passes via a linking "neck" segment into the C-terminal coiled-coil-forming "tail" domain. The main function of the coiled coil is to promote stable dimerization, thereby serving as a stable scaffold to which the actin-binding heads are flexibly attached as they "walk" along the actin filament, although the N-terminal regions of the coiled coils are found to be less rigid to increase the conformational flexibility of the heads (Trybus et al. 1997). However, for the muscle myosins, the coiled coil further mediates lateral association with other myosin molecules into the thick myofilaments (Cohen and Parry 1998). For further details on the fibrous proteins of the micro- and myo-filaments and for their functioning in muscles, the reader is referred to Chaps. 9 (Hitchcock-DeGregori and Barua 2017), 10 (Tskhovrebova and Trinick 2017) and 11 (Squire et al. 2017).

3.2.2 From Static Oligomerization to Dynamic Molecular Recognition

The vast majority of coiled coils of non-filamentous proteins had to wait until the genomic era to be discovered on the basis of their sequence repeats, revealing a vast abundance of coiled coils of all possible lengths. They are especially frequently found as oligomerization motifs in the context of multi-domain proteins, where they specifically recognize and bind to their oligomerization partner(s) by forming a coiled coil. This is the case for most of the filamentous proteins described above, but also for many proteins that are associated with filamentous structures but which do not form filaments themselves. An example is the class of non-muscle myosin motor proteins, which employ their coiled coil only for dimer formation – and to mount cargo molecules to transport them along the actin-microfilament. A similar architecture is also found in the actin-bundling cortexillins. Here, two actin-binding head domains can stably cross-link two actin fibrils, as they are held together by a stable dimeric coiled-coil tail (Faix et al. 1996). Further, the overall architecture of kinesin, one of the two microtubule motor proteins, is analogous to myosin. In kinesin, the head domains are attached via flexible linkers, which are located C-terminally to a dimerizing coiled-coil domain (Fig. 3.2). Similar to myosin, the two heads are thought to walk in a "hand-over-hand" fashion along the microtubule, while cargo is attached via the coiled-coil domain (Asbury et al. 2003; Marx et al. 2009).

A more dynamic existence with changing partners is lived by many transcription factors of the "*b*asic region leucine *zip*per" (b/ZIP) family, which consist of an activation domain, a short coiled-coil dimerization domain, and a basic DNA-binding

domain (Landschulz et al. 1988). As leucine occurs in every seventh position, which corresponds to the heptad position *d*, the coiled-coil domain is referred to as a leucine zipper. The homo-dimeric GCN4 leucine zipper is probably the best studied coiled coil and has served as a model system for many fundamental studies probing the roles of individual heptad positions on stability and specificity (Harbury et al. 1993; Eckert et al. 1998; Liu et al. 2006; Diao 2010). As it is remarkably stable in different oligomeric states, it has been exploited as a stabilizing fusion adaptor in structural studies of different dimeric, trimeric and tetrameric proteins (reviewed in Deiss et al. 2014). However, certain b/ZIP transcription factors can form different homo- and hetero-dimers for further regulation of transcription. One such example is the AP-1 transcription factor, which is usually a heterodimer of the subunits Jun and Fos, but Jun can also form a functional homodimer on its own (Fig. 3.2). Moreover, in proof-of-concept studies, the dimeric association of the AP-1 factor could be abolished by peptides mimicking and competing for the native coiled-coil interactions, which were either selected in an *in vivo* process from a protein library (Mason et al. 2007) or rationally designed (Rao et al. 2013). The dimeric association of certain leucine zipper domains can also be regulated dynamically, as described for the association of c-Myc with its dimer-partner Max, which is most efficient when a histidine residue in the coiled coil is protonated at acidic pH (Lavigne et al. 1995). Again, much of the fine-tuning of stability and partnering specificity is encoded in the nature of the amino acids in - and peripheral to - the core.

Further, an example of an intrinsically highly dynamic coiled-coil oligomerization system is found in the membrane fusion process mediated by the SNARE (for *s*oluble *n*-ethyl maleimide sensitive factor *a*ttachment protein *re*ceptor) proteins, a superfamily with more than 60 members in yeast and mammalian cells. They generally mediate the fusion of vesicles with their target membranes and are best studied in the context of synaptic exocytosis (Zhou et al. 2015). To initiate fusion, three different SNARE proteins attached to two different membranes form a highly stable heterotetrameric coiled coil, the SNARE core-complex. One of the proteins, synaptobrevin, is anchored in the vesicular membrane, exposing one helix with a so-called SNARE motif. The other two proteins are syntaxin (with one) and SNAP-25 (with two SNARE-motif helices), which are both attached to the target membrane, preparing to capture a lone synaptobrevin SNARE-motif helix to fold into the SNARE complex (Fig. 3.2). The zippering-up of this coiled-coil complex is thought to drive the fusion process by progressive juxtaposition of the membranes (Brunger 2005). Once fused, all three SNARE core-complex components find themselves in the same membrane. The oligomerization specificity of the complex is controlled in part by a specific layer of polar residues (Sutton et al. 1998). Further, the formation of the SNARE complex is regulated on several levels, e.g., by the assumption of a self-inhibitory closed helical-bundle conformation of the syntaxin molecule (Munson et al. 2000) or by the binding of specific regulatory factors to syntaxin (Misura et al. 2000). These regulatory interactions are only the tip of the iceberg of an astonishingly complex and dynamic molecular recognition system that is far from fully understood.

3.2.3   Intra-chain Coiled Coils

A non-negligible fraction of all coiled coils are actually not oligomers but are formed between different segments of the same polypeptide chain. Compared to their oligomer-forming relatives, they have a higher tendency to be antiparallel and most of them appear to serve as molecular recognition modules or intramolecular spacers. A classical example for the former is the "helical arm" of the seryl-tRNA synthetase, a coiled coil formed by a long N-terminal α-helical hairpin, which is responsible for the specific recognition of the characteristic elongated variable arm of the corresponding tRNA[Ser] (Chimnaronk et al. 2005). Two further examples of coiled coils of roughly similar functions are described for ClpB and HslU in the section on non-specific molecular recognition of this chapter. A simple example for the functionality as an intramolecular spacer is found in the microtubule motor protein dynein, which is architecturally different and more complex than kinesin. Dynein has separate ATPase and microtubule-binding domains that are connected via a rigid intramolecular anti-parallel coiled-coil spacer (reviewed in Carter et al. 2016).

Especially impressive exemplars of intramolecular spacers are the long antiparallel coiled-coil arms of the SMC (*s*tructural *m*aintenance of *c*hromosomes) proteins. They participate in the higher-order organization of chromosomes and are – with a few exceptions - conserved from bacteria to eukaryotes. While most bacteria have a single SMC protein that forms homodimers, eukaryotes have six different SMC proteins that form three distinct heterodimers that make up the core of three large SMC complexes: the cohesin complex, involved in sister chromatid cohesion, the condensin complex, involved in chromosome condensation, and the less-understood Smc5/6-based complex (Verver et al. 2016). The individual SMC proteins are about 1000-1500 residues long and contain, from N- to C-terminus, the first half of an ATPase domain, the first coiled-coil region, a "hinge" domain, the second coiled-coil region that forms an anti-parallel coiled coil with the first region, and the second half of the ATPase domain. Dimerization is mediated via the globular hinge domain, resulting in V-shaped dimers, wherein the two long arms of the V (the antiparallel coiled coils) are capped by the globular ATPase domains. Via these arms, they embrace chromosomal DNA. For complete enclosure of the DNA, the gap between the two ATPase domains is closed by kleisin, an elongated molecule that additionally mediates interactions with further cofactors (Gligoris et al. 2014; Palecek and Gruber 2015). In EM micrographs, the coiled coils of cohesin and condensin were found to be mostly elongated with a defined kink region in one of the cohesin arms, suggesting only modest flexibility (Anderson et al. 2002). This is in contrast to a recent atomic force microscopy study of condensin, which shows the dimer in a variety of conformations, including V, O, B, and P shapes (Eeftens et al. 2016). In part, the structural flexibility of the SMC proteins can be explained by the antiparallel dimeric nature of their coiled coils. As such, they can bend to higher angles than their higher-oligomeric counterparts, without breaking or unwinding their helices locally. In the following section, we will mostly deal with the converse of this, with coiled-coil spacers that are much more rigid in their folded

state, but which nonetheless exemplify dynamic features that are not apparent from their 3D-structures.

3.3 Molecular Spacers – Dynamic and Resilient

Coiled coils are ubiquitously employed as molecular spacers in fibrous proteins because they provide a simple means of forming elongated structures of defined length. This appeared to be their major role when they were first characterized in protein filaments. However, whereas this role is a rather static one, recent years have revealed a wealth of examples that highlight dynamic properties of coiled-coil spacers. These properties are found on two very different timescales: the classical one concerns protein folding and dynamics in the microsecond regime, based on the biochemical properties of individual coiled-coil sequences. The other one concerns evolutionary dynamics across generations, based on a high degree of resilience towards mutations, including larger insertions or deletions resulting from genetic recombination events. Both properties are represented illustratively in the family of the trimeric autotransporter adhesins and related bacterial and viral surface proteins, which critically depend on both properties.

3.3.1 The Complex Nature of Trimeric Autotransporter Adhesins

Trimeric autotransporter adhesins (TAA) are highly modular, homo-trimeric adhesins found in the outer membrane of gram-negative pathogens (Linke et al. 2006; Qin et al. 2015). They follow a basic head-stalk-anchor architecture from N- to C-terminus; whereas the head is usually thought to mediate adhesion to host cells or components of the extracellular matrix, the stalk projects the head away from its own cell surface, and the anchor embeds the protein in the outer membrane. Simple stalks consist merely of a long trimeric coiled coil, but in more complex TAA, they can contain further head or head-like domains, giving the whole adhesins the appearance of long ropes with many knots that can be several hundred nanometres in length (Fig. 3.3) (Yeo et al. 2004; Szczesny and Lupas 2008; Szczesny et al. 2008; Hartmann et al. 2012; Xiao et al. 2012).

One peculiarity of TAA is their rather complicated biogenesis. TAA constitute the Type 5c secretion system, a subgroup of the Type 5 – the autotransport – secretion system (Fan et al. 2016). As such, the three chains of each TAA are transported to the periplasm via the general secretion system. Once in the periplasm, they need to maintain an unfolded state for subsequent insertion into the outer membrane and export of their passenger domain to the outside. Membrane insertion is mediated via the BAM complex, which inserts four β-strands at the C-terminus of each chain into
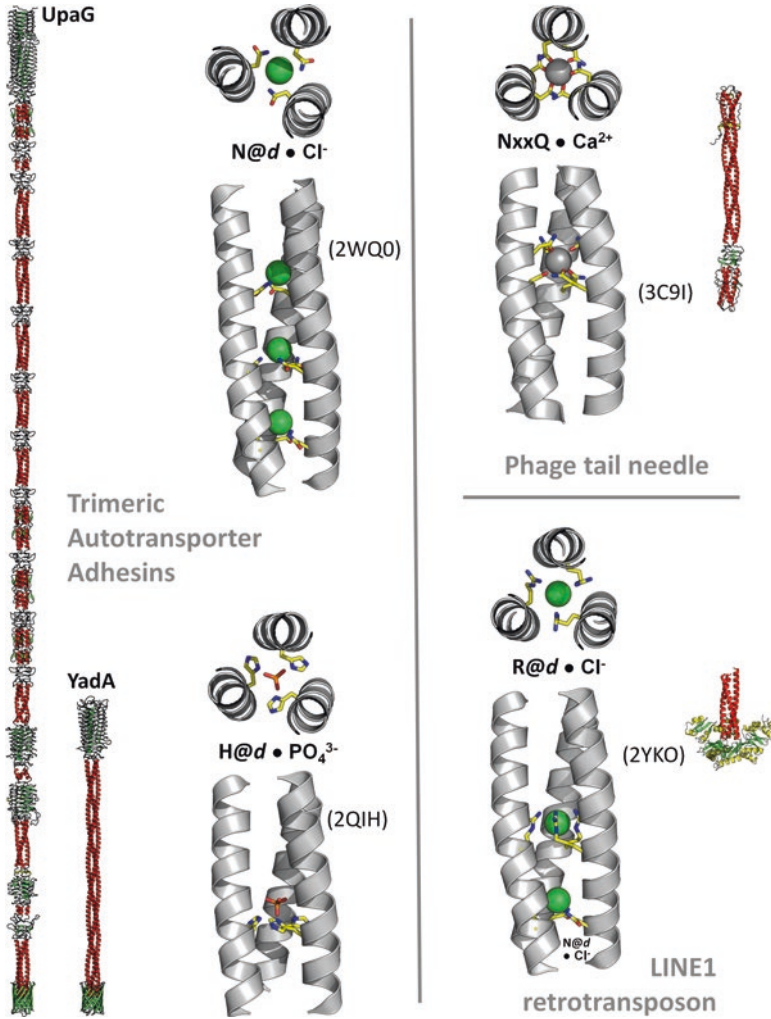
Fig. 3.3 Dynamic coiled-coil spacers with ion-coordinating core motifs. On the *left* side, the full models of the Trimeric Autotransporter Adhesins (TAA) UpaG from uropathogenic *E. coli* and YadA from *Yersinia enterocolitica* (Hartmann et al. 2012). Next to them are close views (*top* and *side*) of two ion-coordinating motifs typically found in TAA stalks. The three consecutive chloride-coordinating N@*d* motifs are from an engineered GCN4 variant, the phosphate-coordinating H@*d* motif is from the stalk of the *Moraxella catarrhalis* TAA UspA1. On the *right* side, the crystal structures of the phage P22 tail-needle GP26 and the retrotransposon LINE-1 ORF1p RNA-packaging protein. The former has a calcium-coordinating motif with asparagine in positions *a* and glutamine in positions *d*, the latter a chloride-coordinating R@*d* motif with glycines in the preceding positions *a*. The depicted ORF1p crystal structure is from a truncated construct and shows only about half of the length of the coiled coil

the outer membrane to form an oligomeric twelve-stranded β-barrel. In a last step, the still unfolded remainder of the protein – the passenger domain – is thought to be exported through the β-barrel to the outside, a process supposedly driven by the progressive folding of the exported domains (Hartmann et al. 2009; Leo et al. 2012).

Another peculiarity of TAA is their highly variable domain architecture. Apart from the anchor, which is a defining element of the family, and a coiled-coil segment that passes into the anchor, there is a large set of conserved domains that can be joined together in many different patterns. While the stalk segments consist mainly of α-helical coiled coils, the head domains, which can occur at the N-terminus but also within the stalk, are mostly of β-stranded architecture. The transitions between the two types, from α to β and from β to α, are mediated by a set of conserved small connector domains. Most of them employ a simplistic structural motif, a *β*-layer, as the shortest way into – and out of – coiled coils, which we will describe at the end of this section (Hartmann et al. 2012, 2016). Interestingly, the high coiled-coil content of TAA was exploited in a number of structural studies. Due to the extreme length of some TAA, only fragments containing a limited number of representative TAA domains were used for structure determination via crystallography (Bassler et al. 2015). These fragments were mostly chosen to start or end in predicted coiled-coil segments, and in many cases, further stabilizing coiled-coil adaptors derived from the GCN4 leucine zipper were fused to their termini (reviewed in Deiss et al. 2014). However, as we will now elucidate, TAA would certainly not exist if their native coiled coils were as rigid as those of the GCN4 adapters.

3.3.2 Polar Core Residues – Between Specificity, Stability, and Foldability

Polar core residues are thought to impart specificity at the cost of lower stability (Eckert et al. 1998), and TAA present an impressive example in which both effects are interwoven. The coiled coils of TAA are conspicuous for a particularly high content of polar and charged core residues, especially in position *d* of the heptad repeat (Hartmann et al. 2009; Bassler et al. 2015). The most frequent residue in this position is asparagine, which leads to the formation of so-called N@*d* layers that coordinate monovalent anions in the centre of the hydrophobic core (Fig. 3.3). They are not specific for a certain type of anion, so the same layer can coordinate monovalent anions of different size and shape – for example chloride, bromide, iodide, or nitrate (Hartmann et al. 2009). Many TAA contain extended runs of consecutive N@*d* layers. Biochemically, the insertion of N@*d* layers was shown to progressively destabilize the structure. When engineered into the GCN4 leucine zipper, a single N@*d* layer led to a dramatic decrease of the melting temperature and a concentration-dependent folding and, with a second and third layer, the constructs only folded at very high concentrations (Hartmann et al. 2009). On the other hand,

however, N@*d* layers were shown to be strong determinants for the heptad register. Attempts to shift the register of a coiled-coil fragment with four consecutive N@*d* layers to an alternative heptad register with a canonical hydrophobic core using out-of-register fusions to stable GCN4 adaptors failed - the coiled coil preferred to incorporate discontinuities at the domain boundaries rather than moving the N@*d* layers out of the *d* positions (Hartmann et al. 2009). With this bipartite effect, lowering stability while conferring register specificity, N@*d* layers appear to fine-tune the folding propensity of the TAA chains, such that they stay in an unfolded but soluble state until their export is complete. Thereafter, the N@*d* layer-containing coiled coils are thought to fold when their local concentration is increased sufficiently by the folding of the adjacent domains.

Apart from N@*d* layers, several other polar core motifs are widely represented in TAA and also in other homo-trimeric coiled coils, such as trimeric accessory factors of TAA (Grin et al. 2014), but also in proteins unrelated to TAA. Albeit, many such motifs do not coordinate ions in the core, but Fig. 3.3 depicts that, in addition to N@*d* layers, three further, especially good examples of central ions have been found. One example is from the phage P22 tail-needle GP26. In addition to N@*d* layers and other polar core residues, it contains a motif with the sequence NxxQD, which coordinates a calcium ion in the core of the bundle with asparagines in positions *a* and glutamines in positions *d* (Olia et al. 2009). The second example is from a stalk segment of the *Moraxella* adhesin UspA1, in which histidine residues are found in different core positions, in one heptad in an *a* position and in two heptads in *d* positions. In two of the three instances, in the *a* and one *d* position, the histidines coordinate a phosphate ion in the core (Conners et al. 2008). In the other *d* position and other known instances of histidines in core positions of trimeric coiled coils, they coordinate a central water molecule, as in the *E. coli* TAA EibD (Leo et al. 2011) or the tail-needle of the first example. Generally, water molecules are not uncommon in the core of trimeric coiled coils with non-canonical core residues (unpublished observation). The last example is from the RNA-packaging protein of the human LINE-1 retrotransposon, which trimerizes via a long, N-terminal coiled coil (Khazina and Weichenrieder 2009). This coiled coil features a GxxR motif, in which arginines in position *d* coordinate a chloride ion in the core at the height of the glycines in the preceding position *a* (Khazina et al. 2011). Interestingly, the whole coiled coil is highly sensitive to point mutations on the functional level (Goodier et al. 2007; Doucet et al. 2010), and mutations of the GxxR motif completely abolish retrotransposon activity (Khazina and Weichenrieder, personal communication). Although in many instances the presence of polar residues in the core is potentially the result of point mutations that are neutral towards structure and function, they appear to generally fine-tune a balance between structural stability and functional dynamics of coiled coils.

3.3.3 Resilience Towards Insertions and Deletions – Boosting Evolvability

With the ability to adopt a continuum of different periodicities apart from the heptad repeat, coiled coils have a remarkable resilience towards insertions and deletions (Hicks et al. 2002). Short insertions and deletions of a few residues can usually be accommodated by local alterations in periodicity (Brown et al. 1996; Lupas and Gruber 2005; Parry et al. 2008). Likewise, insertions or losses of longer segments within a coiled coil, which may result from homologous recombination or duplication events, would usually only lead to local perturbations at the newly formed transitions between segments. Generally, transitions between segments of different periodicity are very common in the coiled coils of natural proteins as is the case, for example, in the dimeric coiled coils of intermediate filament proteins (Chernyatina et al. 2015), trimeric coiled coils of TAA (Alvarez et al. 2010), or the tetrameric coiled coils of tetrabrachion (Stetefeld et al. 2000). Moreover, recent studies systematically investigated the accommodation of small insertions or deletions into coiled coils by a comparative structural analysis of the same coiled coils with and without specific perturbations (Hartmann et al. 2009, 2016). It is generally impressive how easily coiled coils can incorporate almost arbitrary new stretches of sequence, as demonstrated by the crystal structure of an artificial trimeric coiled coil with a sequence spelling the name "Andrei N. Lupas" (with a "v" instead of the "u"), fused to stabilizing GCN4 adaptors. The structure showed how even a proline residue can be accommodated without major structural disturbances (Deiss et al. 2014).

Alternatively, as depicted in Fig. 3.4, insertions that cannot be well accommodated within the coiled coil can be looped-out of the coiled-coil structure. In the simplest case, these include single-residue insertions (skip residues), that – if not accommodated by altering the periodicity – can form Π-turns, which are single larger turns of the α-helix (Lupas 1996). Insertions of more than one residue can be extruded in a similar manner: in bacteriophage T4 fibritin, two short insertions of four and twelve residues are looped-out, leaving the coiled-coil geometry largely unaffected (Tao et al. 1997). Even longer extrusions can be found as conserved motifs or domains in TAA, as in the head of Neisserial adhesin A (NadA) (Malito et al. 2014) or the common FGG domain (Hernandez Alvarez et al. 2008; Edwards et al. 2010). In the latter, the looping-out is accompanied by a 120° rotation of the three chains around the trimer axis, such that the helices N-terminal to the extrusion seem to continue into the helices of the neighbouring chains C-terminal to the extrusion. Generally, in most cases of such extrusions, whether they be long or short, or with or without rotation, the backbone interactions between the alpha helices just before and after the extrusion resemble those of continuous helices. It is as if the helices did not take any notice of the extrusion (Hartmann et al. 2012).

However, there are two kinds of short insertions that almost certainly lead to an interruption of the canonical trimeric coiled-coil structure. The insertion of two or six residues is geometrically too demanding to be accommodated in heptad repeat
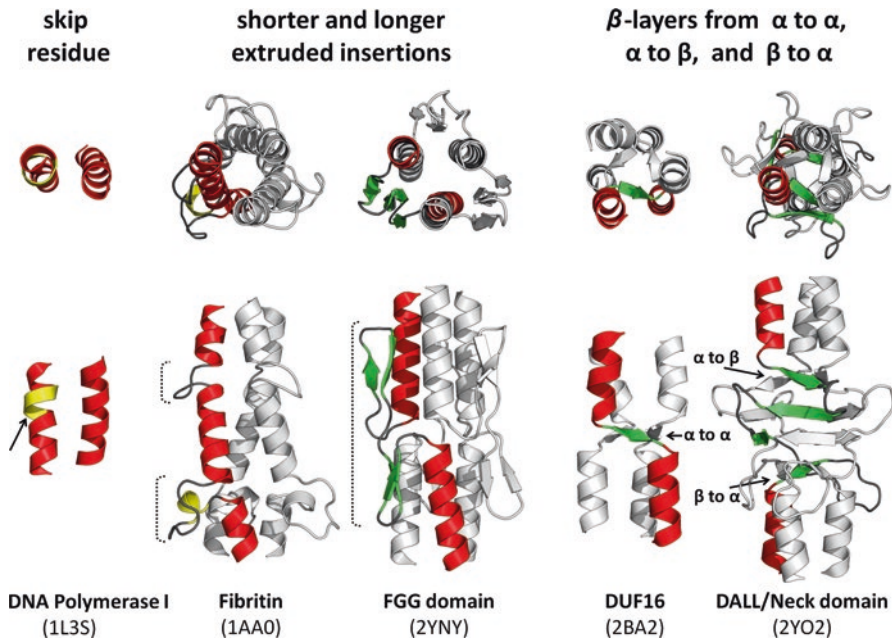
Fig. 3.4 Strategies of coiled coils to cope with insertions. Depicted are insertions in coiled coils that are not accommodated in the form of periodicity changes but extruded from the α-helical coiled-coil structure. Single-residue (skip residue) insertions can lead to the formation of a Π-turn. Longer insertions as in fibritin and in the FGG domain of TAA can be completely looped-out from the coiled coil. A special case are the *β*-layers – they form a β-sheet-like structure in the core of the bundle that can connect two coiled-coil segments but can also form transitions into – and out of – non-α-helical segments

coiled coils and systematically leads to the formation of a supersecondary structure elements termed *β*-layers (Hartmann et al. 2016). They occur after a *c* or *g* position, and substitute for a *d* position by forming a triangular β-sheet-like structure perpendicular to the axis of the trimer. Thereby, they invoke an incomplete 120° rotation of the three chains around the trimer axis, and the coiled coil continues in position *e*. These *β*-layers can be seen as a last missing link in understanding the effect of small insertions into coiled coils. The insertions of one, three, four, five and – trivially – seven residues into the heptad repeat were already covered by the coiled-coil standard model, which can now be extended to a unifying model for α/β coiled coils. This model constitutes the theoretical framework for the resilience of coiled-coil structural integrity towards all possible kinds of mutational events.

Especially in TAA and viral surface fibres, *β*-layers play additional important roles. They are found in multiple conserved connector domains that mediate the transition from coiled coils to *β*-stranded domains and vice versa, forming the actual core of the transition (Hartmann et al. 2012; Bassler et al. 2015; Koiwai et al. 2016). Accordingly, they are very simple universal adaptors to connect *β*-stranded domains to coiled coils, coiled coils to coiled coils, and coiled coils to *β*-stranded domains

(Fig. 3.4). As such, they presumably play key roles in the arms race between these pathogenicity factors and the ever-adapting environment and host defences. The repetitive nature of the coiled coils and *β*-stranded domains of these trimers greatly facilitates the reshuffling of homologous modules by genetic recombination, enabling a rearrangement and fine tuning of binding sites and a shape-shifting to evade detection by the immune system. The many mechanisms of coiled coils to endure harsh mutational stress in the form of almost arbitrary insertions and deletions, together with simplistic ways of forming transitions to other domains, significantly increase the chances of success for such recombination events, while still posing rigid spacers of defined length in the new structural context.

3.4 Non-specific Molecular Recognition

Coiled coils with dynamic roles of a very different kind are found in certain families of chaperones and AAA+ ATPases involved in protein folding und unfolding (Fig. 3.5). Unlike the specific function in the molecular recognition of oligomerization partners described earlier, these coiled coils are involved in the rather non-specific
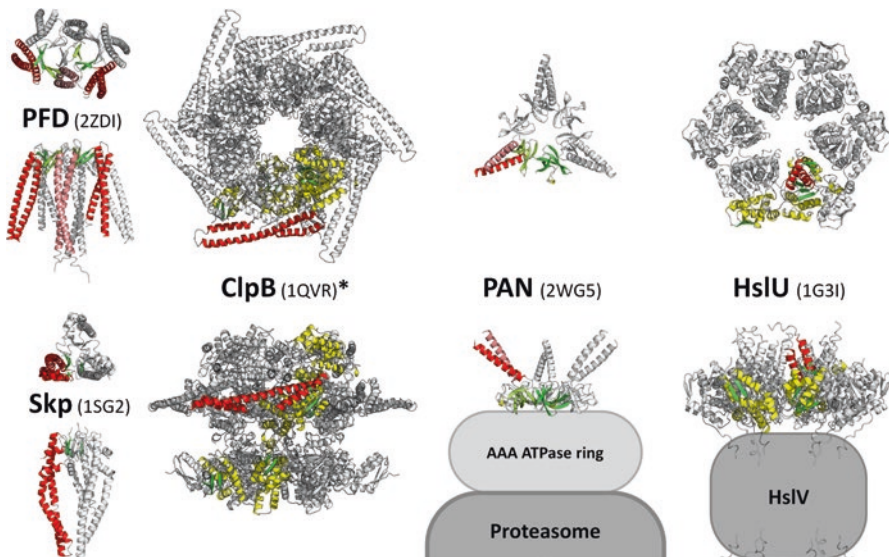
Fig. 3.5 Coiled coils in protein folding and unfolding. One chain of each oligomer is coloured by secondary structure, helices involved in coiled-coil formation in *red*. For PFD, which consists of two different types of subunits, one alpha subunit is in *light*, and two beta subunits in *dark* colour. For the homo-hexameric PAN, two subunits in different conformations are coloured *light* and *dark*. The ATPase domain of PAN, the proteasome, and the HslV domain are indicated as *grey* boxes. The coiled coils of native PAN and of HslU are actually longer than that resolved in the depicted crystal structures. (*) The ClpB ring was modelled on the basis of 1QVR (Martin et al. 2004)

recognition of unfolded and non-native substrate proteins. These coiled coils are predominantly antiparallel and dimeric, and therefore possess a higher degree of intrinsic flexibility than the trimers discussed in the previous section.

3.4.1   Molecular Recognition in Protein Folding and Disaggregation

A particularly well-studied example for coiled-coil mediated non-specific substrate recognition is the hexameric chaperone complex prefoldin (PFD) of archaea and eukaryotes (Vainberg et al. 1998). It can detect, bind, and stabilize nascent chains and deliver them to the archaeal thermosome or the eukaryotic *c*haperonin-*c*ontaining *T*CP1 (CCT) chaperonins for the completion of folding. In eukaryotes, this is of special importance for the folding of actin and tubulin. PFD is assembled from two α and four β subunits. While archaea encode only one or two α subunits and only a single β subunit, the six subunits of eukaryotes are fully differentiated (Leroux et al. 1999). PFD has a jellyfish-like appearance: it has six long anti-parallel coiled-coil tentacles protruding from a main body composed of two β-barrels (Siegert et al. 2000). Each subunit starts and ends with helices that associate into the dimeric tentacles, so each coiled coil is contributed by another subunit.

The tips of these coiled coils, that is, the N- and C-terminal ends, are the major substrate recognition sites of PFD (Siegert et al. 2000). Due to the overall length of the coiled coils, their distal ends have a certain conformational flexibility. They are found in different orientations in different crystal structures and can presumably sample a wide range of bending angles (Ohtaki et al. 2008). Moreover, towards their tips, their dimeric interface appears to become weaker. Where not constrained by crystal contacts, the N- and C-terminal ends of the coiled coils are frayed out and are not resolved in most crystal structures. Therefore, the distal ends do not have a fully defined, rigid structure and can therefore expose different amphiphilic surfaces at different angles. This structural flexibility is implied to be a key in the recognition of hydrophobic surfaces of non-native substrate proteins (Martin et al. 2004; Ohtaki et al. 2008). The precise sequence of the distal coiled coil is not decisive; mutants of archaeal PFD, in which the tips of the N- and C-terminal helices of the tentacles were swapped, still showed activity, as did mutants in which the tips were exchanged for those of eukaryotic PFD (Lundin et al. 2004). The same mode of action is employed by a smaller bacterial PFD-analogue, the periplasmic *s*eventeen-*k*ilodalton *p*rotein (Skp). Skp stabilizes the unstructured chains of outer membrane proteins as they enter the periplasm through the Sec machinery, until they are assembled and integrated into the outer membrane. It has an overall jellyfish-like architecture similar to PFD, albeit with just three anti-parallel coiled coils belonging to three subunits. These subunits, however, do not start and end with the helices of the coiled coil. They start and end in the central beta barrel-like main body, from

which the anti-parallel coiled coils protrude as long helical hairpins (Korndorfer et al. 2004; Walton and Sousa 2004).

A yet different dynamic role was ascribed to the coiled coils of the AAA+ chaperone *c*aseino*l*ytic *p*rotease component *B* (ClpB). This hexameric AAA+ ATPase, involved in the refolding of aggregated proteins, has six long solvent-exposed anti-parallel coiled coils arranged like the blades of a propeller around its perimeter (Lee et al. 2004). These coiled coils, constituting the so-called middle domain, were shown to be able to undergo hinged rigid body motions. As the disaggregation activity of ClpB could be suppressed by fixing the coiled coils in a conformation close to the body of the ATPase via disulphide crosslinks, the coiled coils had been implied in substrate binding and disaggregation (Lee et al. 2003). However, the reactivation of aggregated proteins by ClpB depends on complex formation with Hsp70 (Haslberger et al. 2007), and a recent study has shown that these coiled coils are direct mediators of this interaction (Carroni et al. 2014). Therefore, the coiled coils of the ClpB middle domain potentially play roles in both substrate recognition and higher-order complex formation, a feature that they might have in common with the unfolding machines illuminated in the following section.

3.4.2   Molecular Recognition in Protein Unfolding and Degradation

Another context in which coiled coils were implied in the recognition of non-native proteins is the proteasome system, which is found in archaea, eukaryotes, and the actinobacterial lineage. The proteasome is formed by four stacked rings of seven subunits each, which confines its proteolytic compartment to the inside of a narrow cylinder. Access to the inside is regulated by the gatekeeping ATPases, capping the proteasome on both ends, which recognize and unfold substrate proteins to feed them into the proteasome in an ATP-dependent manner. The main gatekeepers are the archaeal homo-hexameric *p*roteasome *a*ctivating *n*ucleotidase (PAN), the eukaryotic hetero-hexameric homolog *r*egulatory *p*article *t*riple-A ATPase (RPT), and the actinobacterial homo-hexameric homolog *A*TPase forming *r*ing-shaped *c*omplexes (ARC). These hexameric rings can be subdivided into a ring of N-terminal substrate recognition domains that is stacked on a ring of the C-terminal ATPase domains; the latter forms the interface to the proteasome and provides the energy for substrate unfolding and translocation. While there are multiple other ATPases with other substrate recognition domains that can act as proteasomal gatekeepers (Forouzan et al. 2012), a very unique feature of the PAN, RPT, and ARC N-domains is their long N-terminal dimeric coiled coils, which protrude from a ring of OB domains (Djuranovic et al. 2009; Zhang et al. 2009).

Structurally, these N-terminal coiled coils break the sixfold symmetry of the hexamer down to threefold symmetry: each subunit encodes only one helix of the dimeric coiled coil, so that the hexamer is armed with only three protruding coiled

coils. Thus, the hexameric ATPase can be seen as a homo-trimer of asymmetric dimers, whereas the conformation of the two monomers in each dimer differs merely in the cis/trans isomeration of a double-proline linker at the base of the coiled coil, controlling the orientation of the N-terminal helix and thus specifying the partnering in coiled-coil formation. Functionally, these coiled coils are necessary for an energy-independent chaperone activity of the ATPases. This chaperone activity could be ascribed to the hexameric N-domains alone: different constructs based on the PAN and ARC N-domains could prevent the aggregation of heat-denatured proteins in vitro, as long as they included both coiled coils and an OB domain with an intact double-proline linker. Interestingly, the nature of the coiled coil was not decisive, as chimeric constructs with the GCN4 derived constructs impaired similar chaperone activity (Djuranovic et al. 2009). Similar to the PFD tentacles, this mode of action is supposedly based on a basal ability of coiled coils to display a continuum of amphiphilic interaction surfaces at (and towards) their poorly structured, unconstrained termini.

It should be noted, however, that the eukaryotic RPT hexamer also forms the base of the large 19S regulatory particle. On to this base, the multisubunit lid of the 19S particle is mounted, whereas many of the protein-protein interactions are mediated via the RPT coiled coils (Lander et al. 2012). Here, the coiled coils are not involved in initial substrate recognition but rather in the assembly of the complex.

Furthermore, most bacteria possess a functional analogue to the proteasome system. It is based on the protease HslV, which is composed of two stacked hexameric rings confining the proteolytic centre in the inside, and which is regulated by the hexameric HslU ATPase (Sousa et al. 2000). Similar to the described proteasomal gatekeepers, HslU employs coiled coils for substrate recognition, which are located in their I-domains. These domains protrude around the central pore, one per subunit, and resemble long hairpins. Structurally, these hairpins are supposed to consist mostly of antiparallel coiled-coil segments. Most of their structure is, however, unresolved in crystal structures of functional HslU:HslV complexes, indicating a high level of flexibility, and their deletion impairs substrate recognition (Song et al. 2000). Hence, the non-specific recognition of non-native proteins by coiled coils is a reoccurring theme in protein folding and unfolding, in chaperones and in the gatekeepers of proteolytic machines.

3.5   Coiled Coils in Signal Transduction

Over the last decade, a wealth of experimental evidence has accumulated on how signals may be dynamically transmitted along the length of coiled coils. In the previous sections, we covered dynamic aspects determining the folding and oligomerization of coiled coils, as well as aspects of molecular recognition governing the interplay of coiled coils with other biomolecules. However, the dynamics of signal

transduction along coiled coils seem to operate on yet another level, based on changes in the coiled-coil packing mode. A particularly well studied model system for these mechanisms is the family of receptor histidine kinases. These are proteins of especially complex and interesting coiled-coil architecture, with different conserved segments of parallel and antiparallel, dimeric and tetrameric coiled coils, passing into each other to form an elongated signal relay chain (Fig. 3.6). They are the first component of bacterial two-component signal transduction systems (Bhate et al. 2015), where they act as a sensor of the extracellular environment. Signals generated by the extracellular domain control the autophosphorylation of a cytoplasmic histidine. Subsequently, the phosphoryl group is transferred to the second component of the system, the response regulator, which modulates its interactions with downstream targets, most commonly as a transcription factor. Even though the precise mechanism of signal transduction along the coiled coils of receptor histidine kinases is still a field of heavy debate, there has been tremendous progress in recent years.
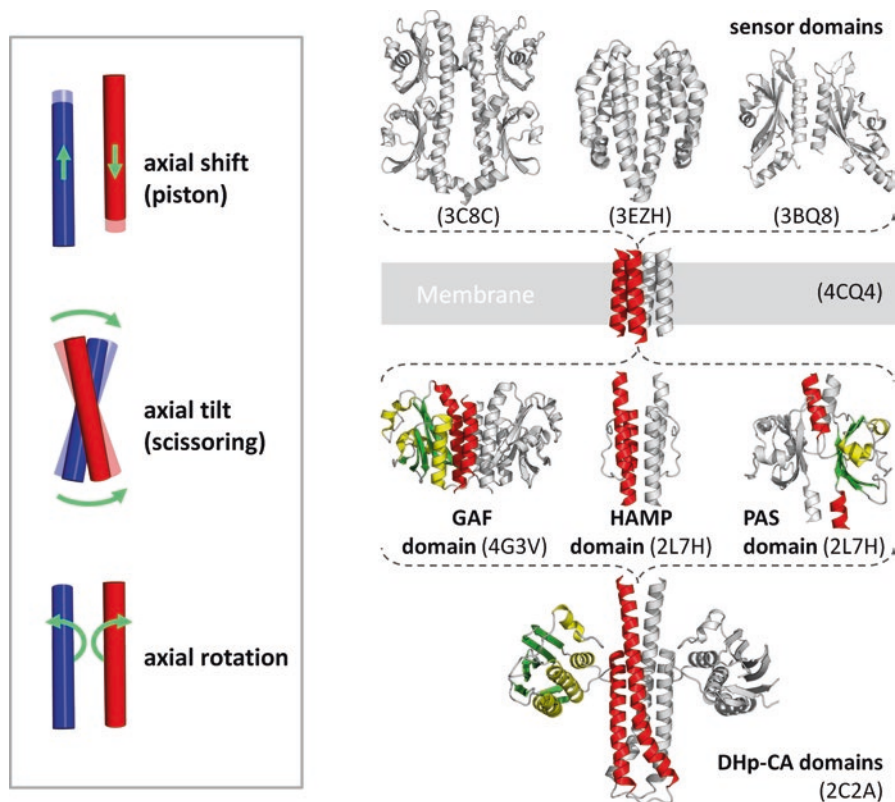
Fig. 3.6 Coiled coils in signal transduction. The box on the *left* shows the three classical models for coiled-coil signal transduction. The scheme on the *right* illustrates the modular architecture of transmembrane histidine kinases

3.5.1 Basic Architecture and Signal Propagation in Histidine Kinases

Although the domain architecture of histidine kinases is highly variable across the members of the family, they possess a set of conserved helical regions and coiled-coil domains that are crucial for oligomerization, signalling, and regulation. Histidine kinases are typically membrane-bound, and the simplest representatives consist only of an extracytoplasmic sensor domain, a transmembrane region, and the intracellular catalytic machinery. The transmembrane region is usually an anti-parallel four-helical bundle, with two antiparallel helices from each protomer (Gordeliy et al. 2002; Hartmann et al. 2014). The first helix is usually at the extreme N-terminus and leads to the outside, passing into the sensor domain. The sensor domains can be of different folds, but all of them end in an α-helix – termed the p-helix (periplasmic helix) – that passes into the second transmembrane helix (Bhate et al. 2015). Deviations from this rule can be found in histidine kinases with membrane-integral sensors (Saita et al. 2015) or with sensor domains preceding the first transmembrane helix (Garcia et al. 2016). In any case, after traversing the membrane, the second transmembrane helices continue into the catalytic machinery, either via short helical linkers or via additional signal transducing domains. The first domain of the catalytic machinery is the *d*imerization and *h*istidine *p*hosphotransfer (DHp) domain, an antiparallel four-stranded coiled coil consisting of two antiparallel helices from each protomer. It contains the catalytic histidine and is followed by the flexibly attached C-terminal *ca*talytic (CA) domain that harbours the ATP binding site. For catalysis, the CA domain is docked to the membrane-proximal part of the DHp domain (Bhate et al. 2015).

To invoke activity, signals produced by the sensor domain have to reach the DHp/CA domains. It is well-accepted in the field that the signal is output from the sensor domain via its C-terminal p-helices that pass into the second transmembrane helices. From there, depending on the architecture of a particular histidine kinase, these signals might need to pass additional signal transmission domains before reaching the DHp. However, the core of these additional domains is either formed of a tetrameric coiled coil, as in GAF domains (Batchelor et al. 2013) or HAMP domains (Hulko et al. 2006), or these domains start and end with dimeric coiled-coil segments as in PAS domains (Wang et al. 2013). In any case, for signal propagation towards the DHp, the signal received via the N-terminal helices of these domains must be transmitted to their C-terminal helices, so the signal within any histidine kinase traverses along coiled coils and coiled coil-like helical bundles.

The three classical models for signal propagation in helical bundles are based on different degrees of freedom of the constituent helices, which are treated as rather rigid entities: the piston model, the scissor model, and the rotation model, in broad chronological order. In the translation-based piston model, ligand binding is thought to induce a ~1 Å piston motion of the p-helices, orthogonal to the membrane (Chervitz and Falke 1996; Falke and Erbse 2009), which is thought to be passed on through the transmembrane helices to the inside. This model would, however, only

lead to coiled-coil packing changes in cases where the individual helices undergo piston motions with respect to each other. In the scissor model, the p-helices are thought to undergo a symmetrical scissor-motion, a rotational motion orthogonal to the bundle axis, widening the bundle at the interface to the membrane (Milburn et al. 1991; Lowe et al. 2012; Molnar et al. 2014). Thereby, the signal could propagate in a symmetric fashion via changes in the sheering angle of the helices. This is effectively a modulation of the coiled-coil crossing angle and thus leads to a distortion of the core packing, simulating a change in periodicity and sequence repeat length. In the third classical model, helix rotation, the helices are thought to undergo concerted rotations parallel to the bundle axis (Hulko et al. 2006; Neiditch et al. 2006). Here, the rotation angle of the individual helices would directly influence the packing geometry in the core of the bundle, by rotating certain positions further into – or further out of – the core of the bundle. In contrast to the scissoring model, in which the length of a single scissoring segment would be physically limited, the rotation model has no obvious length limitation. It could therefore rationalize how a signal could be transduced along the full length of a multi-domain histidine kinase. This model celebrated much attention when the first structure of a HAMP domain was solved.

3.5.2 Signal Propagation Through HAMP Domains

The HAMP domain is named after its occurrence in *h*istidine kinases, *a*denylyl cyclases, *m*ethyl-accepting chemotaxis proteins, and *p*hosphatases (Aravind and Ponting 1999). It was shown that HAMP domains from different proteins and contexts can be interchanged to yield functional chimeras, as between histidine kinases and chemotaxis receptors (Utsumi et al. 1989; Zhu and Inouye 2003), between different histidine kinases (Appleman et al. 2003), or between chemotaxis receptors and adenylyl cyclases (Kanchan et al. 2010), implying that HAMP domains have a common, general mechanism. This particular mechanism appeared to be engraved in the HAMP structure that Hulko and colleagues determined in 2006 from the putative archaeal receptor Af1503 (Hulko et al. 2006). The structure showed a symmetric homodimer forming a parallel four-helical coiled coil, in which the N-terminal helices are connected by an extended linker to the C-terminal helices. The striking feature of this structure resides in the core packing of the bundle. It does not assume classical "knobs-into-holes" but a special case of "knobs-to-knobs" packing. This packing mode is generally found in coiled coils deviating from the heptad repeat, where it forms *x*- and *da*-layers: in *x*-layers, the core residues point directly towards the coiled-coil axis in an acute geometry, while in *da*-layers, two consecutive core residues point sideways in opposite directions, forming a ring of interacting residues around the bundle. However, in the HAMP domain structure, the N- and C-terminal helices are staggered by one helical turn, resulting in a mixed, "complementary *x*-*da*" packing: when the N-helices are in *x*-geometry, the C-helices are in *da*-geometry in the same layer, and vice versa (see also Chap. 4 (Lupas et al. 2017)).

The complementary *x-da* packing of HAMP could potentially be interconverted into classical knobs-into-holes packing via a concerted axial rotation of all four helices. This resulted in the gearbox model, in which the N- and C-terminal helices of the bundle rotate about 26° like cogwheels in opposite directions (Hulko et al. 2006). Thereby, HAMP domains in histidine kinases would translate a rotational signal received through the transmembrane helices into a rotational signal of opposite sense and propagate it to the DHp domain. Although the expected alternative, fully-rotated *knobs-into-holes* state of Af1503 HAMP has not been observed, different intermediate states could be brought about by point mutations in one specific core position. In chimeric constructs, the rotational state of these HAMP mutants correlated with the activity of the downstream adenylyl cyclase or chemotaxis receptor output modules (Hulko et al. 2006; Ferris et al. 2011).

Further structural studies of Af1503 HAMP fusion constructs gave insight into possible mechanisms of how a rotational signal might translate into the activation of downstream domains. A NMR study of a fusion to a DHp domain suggested that a rotational input signal is attenuated over the length of the DHp domain. A rotation would thus be strongest at the membrane-proximal side, which is involved in the positioning of the catalytic domains, and less influential further downwards, at the position of the catalytic histidine and the binding region of the response regulator. This implied that the regulation of histidine kinases via the rotation model would be mostly based on the positioning of the catalytic domains rather than the positioning of the histidine (Ferris et al. 2012), which will be discussed in more detail at the end of this section. Also, methyl-accepting chemoreceptors could be regulated by a similar mechanism. These proteins are, very much like histidine kinases, extended homodimeric coiled-coil bundles, whose cytosolic portion generally comprises a HAMP domain and a long methyl-accepting domain (Aravind and Ponting 1999; Alexander and Zhulin 2007). The actual signalling behaviour of these chemoreceptors is fine-tuned by the methylation state of specific glutamate and glutamine residues in the latter domain. A recent study using a fusion of the AF1503 HAMP to a chemoreceptor methyl-accepting domain suggested that different rotational states might modulate the affinity for the methylation/demethylation system (Ferris et al. 2014b). In turn, different methylation states might tune the traversal of signals towards the signalling subdomain.

The gearbox model, however, has not remained uncontested. With more available structures of HAMP domains, a fairly large conformational variation between different HAMP domains became apparent (Dunin-Horkawicz and Lupas 2010a, b). In addition to the full range of axial rotation states – from complementary *x-da* to canonical *knobs-into-holes* – different structures of different HAMPs show different crossing angles and also different axial shifts of the helices. Thus, depending on which two individual structures are compared, one could conclude that either axial rotation, scissoring motions, piston motions, or all possible combinations thereof are involved. However, as long as such comparisons are not based on two states of the same HAMP, but on two different HAMP domains, conclusions are to be drawn with care. A scissoring mechanism was proposed on the basis of the structure of a tandem HAMP repeat (Airola et al. 2010) and has further experimental support in

crosslinking studies (Swain and Falke 2007; Swain et al. 2009; Watts et al. 2011). More recently, similar studies implied a combined scissoring and rotation mechanism for histidine kinases (Matamouros et al. 2015). Also, the piston model remains under active discussion (Hazelbauer et al. 2008; Ames et al. 2016) and is the subject of molecular dynamics simulations (Gushchin et al. 2013; Zhu et al. 2013).

All the different classical signalling mechanisms have one commonality – they are based on the switching between two conformationally-distinct structural states (Bhate et al. 2015; Schultz et al. 2015). This is not strictly the case for a rather recent alternative two-state model of signal transduction, which is of a different nature and cannot be derived from individual crystal structures: the dynamic bundle model. Therein, the signalling state is not necessarily of different structure than the "off" state, but possesses a higher degree of flexibility or disorder (Parkinson 2010; Zhou et al. 2011; Ames et al. 2014). Since such a more dynamic state would generally allow for more conformational freedom, this model is essentially compatible with all classical models of axial rotation, axial tilt (scissoring), and axial shift (piston) of the helices. In the "off" state, the bundle would rest in a rigid state that imposes conformational restraints on the downstream domains. In the dynamic "on" state, the conformational restraints are released, allowing the downstream domains to carry out their activities. The idea of the dynamic bundle culminates in the concept of signalling via "regulated unfolding", in which the domain would partially or fully unfold to release conformational restraints (Schultz and Natarajan 2013).

3.5.3 The Role of Asymmetry

Independent of the exact mechanism of signal propagation along coiled coils, there is increasing evidence for an asymmetric component to the auto-phosphorylation mechanism of histidine kinases. The DHp domain is generally connected via continuous helices to its upstream domain. However, in essentially all available crystal structures, these connecting helices are found in asymmetric conformations – at least in structures where they occur after upstream HAMP, PAS, or LOV domains (Diensthuber et al. 2013; Wang et al. 2013; Ferris et al. 2014a; Mechaly et al. 2014) – and these asymmetric transitions seem to play a role in the positioning of the CA domains. Generally, the connecting helices resemble a loosely-packed dimeric coiled coil with unusual properties. In examples between HAMP and PAS, between PAS and DHp, and between HAMP and DHp, the helical connectors have a wider diameter than expected for dimeric coiled coils, and have a high content in polar and charged core residues (Wang et al. 2013; Ferris et al. 2014a). Moreover, they usually deviate from heptad periodicity by having one or more insertions of four residues (stutters) (Ferris et al. 2012). Also, this stutter often coincides with a conserved charged sequence motif and is found in different asymmetric conformations in different crystal structures (Ferris et al. 2012).

Although multiple structures of DHp-CA domain tandems have been determined, so far only two of them have been proposed to be in a potentially active

auto-kinase conformation (Wang et al. 2013; Ferris et al. 2014a). In both cases, the docking of one CA domain in the respective catalytic conformation involves an asymmetric kink or bend of the connecting helices. Although the two putative active conformations differ from each other, both imply that only one CA domain can dock in the given conformation at a time – the other CA domain would remain in an inactive conformation until the release of the first one, so the bundle could kink or bend to the other side, enabling the docking of the other domain. This model is referred to as the sequential flip-flop autokinase mechanism (Ferris et al. 2014a). It is backed-up by the observation of a fast accumulation of the hemi-phosphorylated state under ATP-excess conditions (Casino et al. 2014), and, more generally, by cysteine scanning experiments showing a correlation between conformational flexibility of the connecting helices and kinase activity (Lesne et al. 2016).

A key to this flexibility lies supposedly in the uncommon coiled-coil features of the connecting helices. The loose packing of the largely hydrophilic core, coinciding with a change in the periodicity and register due to insertions into the heptad repeat, seems to predetermine a breaking point of symmetry to allow for an asymmetric CA docking in the asymmetric sequential flip-flop auto-kinase mechanism. But how could these asymmetric transitions be regulated? It seems unlikely that the directionality of the asymmetric transition could be explicitly set or reversed in response to received signals. It is, however, plausible that the general tendency to kink and bend can be influenced by the tightening and loosening of conformational restraints, as set and received from upstream domains. Be it different rotational states, shearing states, or folding/unfolding transitions of the upstream domains that impose or release an initial tension on the two connecting helices, all could potentially be unified in a general dynamic model that knows two states: flexibility for activity in the "on"-, and rigidity in the "off"- state.

3.6 Concluding Remarks

The coiled-coil field has long come of age. However, as an "open" field that is interlacing the whole protein universe, it is steadily growing and has reached a level of complexity that was certainly unforeseen in the early days, when coiled coils were mostly seen as permanent fibrous complexes of impressive stability and tensile strength. This chapter was aimed at depicting the field from its beginnings, while staying abreast of the developments, but with a strong emphasis on the insights into new dynamic properties gained in the last decade. In doing so, many compromises had to be made, and many interesting coiled-coil proteins that had been described in previous reviews fell by the wayside. These are, for example, the pH-dependent shape-shifting coiled coils of haemagglutinin and the hyperstable tetramer tetrabrachion (Burkhard et al. 2001), larger assemblies like the coiled coil-like barrels formed in the outer membrane protein TolC or in bacteriophage coat proteins, described by Lupas and Gruber (2005), or impressive coiled-coil design successes (Woolfson et al. 2012; Tebo and Pecoraro 2015). However, redundancy is also

increasing with complexity all across the functional spectrum, also between coiled coils in completely unrelated biological processes. The more we learn about the functional and structural roles of coiled coils as we find and dissect them in yet new contexts, the more it becomes apparent how astonishingly well we can already understand and describe their functioning at the amino-acid sequence level. It is not easy to imagine how many more surprises are still to come, and what they might be. Until now, however, they have come readily and steadily.

References

Airola MV, Watts KJ, Bilwes AM, Crane BR (2010) Structure of concatenated HAMP domains provides a mechanism for signal transduction. Structure 18:436–448

Alexander RP, Zhulin IB (2007) Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proc Natl Acad Sci U S A 104:2885–2890

Alvarez BH, Gruber M, Ursinus A, Dunin-Horkawicz S, Lupas AN, Zeth K (2010) A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins. J Struct Biol 170:236–245

Ames P, Zhou Q, Parkinson JS (2014) HAMP domain structural determinants for signalling and sensory adaptation in Tsr, the Escherichia coli serine chemoreceptor. Mol Microbiol 91:875–886

Ames P, Hunter S, Parkinson JS (2016) Evidence for a Helix-Clutch mechanism of transmembrane signaling in a bacterial chemoreceptor. J Mol Biol 428:3776–3788

Anderson DE, Losada A, Erickson HP, Hirano T (2002) Condensin and cohesin display different arm conformations with characteristic hinge angles. J Cell Biol 156:419–424

Appleman JA, Chen LL, Stewart V (2003) Probing conservation of HAMP linker structure and signal transduction mechanism through analysis of hybrid sensor kinases. J Bacteriol 185:4872–4882

Aravind L, Ponting CP (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett 176:111–116

Asbury CL, Fehr AN, Block SM (2003) Kinesin moves by an asymmetric hand-over-hand mechanism. Science 302:2130–2134

Bassler, J., Hernandez Alvarez, B. Hartmann, M. D. & Lupas, A. N. 2015. A domain dictionary of trimeric autotransporter adhesins. Int J Med Microbiol, 305:265–275

Batchelor JD, Lee PS, Wang AC, Doucleff M, Wemmer DE (2013) Structural mechanism of GAF-regulated sigma(54) activators from Aquifex aeolicus. J Mol Biol 425:156–170

Bhate MP, Molnar KS, Goulian M, Degrado WF (2015) Signal transduction in histidine kinases: insights from new structures. Structure 23:981–994

Bouameur J-E, Magin TM (2017) Lessons from animal models of cytoplasmic intermediate filament proteins. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Brown JH, Cohen C, Parry DAD (1996) Heptad breaks in alpha-helical coiled coils: stutters and stammers. Proteins 26:134–145

Brunger AT (2005) Structure and function of SNARE and SNARE-interacting proteins. Q Rev Biophys 38:1–47

Burkhard P, Stetefeld J, Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. Trends Cell Biol 11:82–88

Carroni M, Kummer E, Oguchi Y, Wendler P, Clare DK, Sinning I, Kopp J, Mogk A, Bukau B, Saibil HR (2014) Head-to-tail interactions of the coiled-coil domains regulate ClpB activity and cooperation with Hsp70 in protein disaggregation. Elife 3:e02481

Carter AP, Diamant AG, Urnavicius L (2016) How dynein and dynactin transport cargos: a structural perspective. Curr Opin Struct Biol 37:62–70

Casino P, Miguel-Romero L, Marina A (2014) Visualizing autophosphorylation in histidine kinases. Nat Commun 5:3258

Chernyatina AA, Guzenko D, Strelkov SV (2015) Intermediate filament structure: the bottom-up approach. Curr Opin Cell Biol 32:65–72

Chervitz SA, Falke JJ (1996) Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. Proc Natl Acad Sci U S A 93:2545–2550

Chimnaronk S, Gravers Jeppesen M, Suzuki T, Nyborg J, Watanabe K (2005) Dual-mode recognition of noncanonical tRNAs(Ser) by seryl-tRNA synthetase in mammalian mitochondria. EMBO J 24:3369–3379

Cohen C, Parry DAD (1998) A conserved C-terminal assembly region in paramyosin and myosin rods. J Struct Biol 122:180–187

Conners R, Hill DJ, Borodina E, Agnew C, Daniell SJ, Burton NM, Sessions RB, Clarke AR, Catto LE, Lammie D, Wess T, Brady RL, Virji M (2008) The Moraxella adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. EMBO J 27:1779–1789

Deiss S, Hernandez Alvarez B, Bar K, Ewers CP, Coles M, Albrecht R, Hartmann MD (2014) Your personalized protein structure: Andrei N. Lupas fused to GCN4 adaptors. J Struct Biol 186:380–385

Diao J (2010) Crystal structure of a super leucine zipper, an extended two-stranded super long coiled coil. Protein Sci 19:319–326

Diensthuber RP, Bommer M, Gleichmann T, Moglich A (2013) Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators. Structure 21:1127–1136

Djuranovic S, Hartmann MD, Habeck M, Ursinus A, Zwickl P, Martin J, Lupas AN, Zeth K (2009) Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. Mol Cell 34:580–590

Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N (2010) Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet 6:e1001150

Dunin-Horkawicz S, Lupas AN (2010a) Comprehensive analysis of HAMP domains: implications for transmembrane signal transduction. J Mol Biol 397:1156–1174

Dunin-Horkawicz S, Lupas AN (2010b) Measuring the conformational space of square four-helical bundles with the program samCC. J Struct Biol 170:226–235

Eckert DM, Malashkevich VN, Kim PS (1998) Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. J Mol Biol 284:859–865

Edwards TE, Phan I, Abendroth J, Dieterich SH, Masoudi A, Guo W, Hewitt SN, Kelley A, Leibly D, Brittnacher MJ, Staker BL, Miller SI, Van Voorhis WC, Myler PJ, Stewart LJ (2010) Structure of a Burkholderia pseudomallei trimeric autotransporter adhesin head. PLoS One 5:e12803

Eeftens JM, Katan AJ, Kschonsak M, Hassler M, De Wilde L, Dief EM, Haering CH, Dekker C (2016) Condensin Smc2-Smc4 dimers are flexible and dynamic. Cell Rep 14:1813–1818

Faix J, Steinmetz M, Boves H, Kammerer RA, Lottspeich F, Mintert U, Murphy J, Stock A, Aebi U, Gerisch G (1996) Cortexillins, major determinants of cell shape and size, are actin-bundling proteins with a parallel coiled-coil tail. Cell 86:631–642

Falke JJ, Erbse AH (2009) The piston rises again. Structure 17:1149–1151

Fan E, Chauhan N, Udatha DB, Leo JC, Linke D (2016) Type V secretion systems in bacteria. Microbiol Spectr 4:VMBF-0009–2015

Ferris HU, Dunin-Horkawicz S, Mondejar LG, Hulko M, Hantke K, Martin J, Schultz JE, Zeth K, Lupas AN, Coles M (2011) The mechanisms of HAMP-mediated signaling in transmembrane receptors. Structure 19:378–385

Ferris HU, Dunin-Horkawicz S, Hornig N, Hulko M, Martin J, Schultz JE, Zeth K, Lupas AN, Coles M (2012) Mechanism of regulation of receptor histidine kinases. Structure 20:56–66

Ferris HU, Coles M, Lupas AN, Hartmann MD (2014a) Crystallographic snapshot of the Escherichia coli EnvZ histidine kinase in an active conformation. J Struct Biol 186:376–379

Ferris HU, Zeth K, Hulko M, Dunin-Horkawicz S, Lupas AN (2014b) Axial helix rotation as a mechanism for signal regulation inferred from the crystallographic analysis of the E. coli serine chemoreceptor. J Struct Biol 186:349–356

Forouzan D, Ammelburg M, Hobel CF, Stroh LJ, Sessler N, Martin J, Lupas AN (2012) The archaeal proteasome is regulated by a network of AAA ATPases. J Biol Chem 287:39254–39262

Fraser RDB, Parry DAD (2017) Structural transition of trichocyte keratin intermediate filaments during development in the hair follicle. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Garcia D, Watts KJ, Johnson MS, Taylor BL (2016) Delineating PAS-HAMP interaction surfaces and signalling-associated changes in the aerotaxis receptor Aer. Mol Microbiol 100:156–172

Gligoris TG, Scheinost JC, Burmann F, Petela N, Chan KL, Uluocak P, Beckouet F, Gruber S, Nasmyth K, Lowe J (2014) Closing the cohesin ring: structure and function of its Smc3-kleisin interface. Science 346:963–966

Goodier JL, Zhang L, Vetter MR, Kazazian HHJR (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol Cell Biol 27:6469–6483

Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, Schlesinger R, Buldt G, Savopol T, Scheidig AJ, Klare JP, Engelhard M (2002) Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. Nature 419:484–487

Grin I, Hartmann MD, Sauer G, Hernandez Alvarez B, Schutz M, Wagner S, Madlung J, Macek B, Felipe-Lopez A, Hensel M, Lupas A, Linke D (2014) A trimeric lipoprotein assists in trimeric autotransporter biogenesis in enterobacteria. J Biol Chem 289:7388–7398

Guharoy M, Szabo B, Contreras Martos S, Kosol S, Tompa P (2013) Intrinsic structural disorder in cytoskeletal proteins. Cytoskeleton (Hoboken) 70:550–571

Gunning PW, Hardeman EC, Lappalainen P, Mulvihill DP (2015) Tropomyosin – master regulator of actin filament function in the cytoskeleton. J Cell Sci 128:2965–2974

Gushchin I, Gordeliy V, Grudinin S (2013) Two distinct states of the HAMP domain from sensory rhodopsin transducer observed in unbiased molecular dynamics simulations. PLoS One 8:e66917

Guzenko D, Chernyatina AA, Strelkov SV (2017) Crystallographic studies of intermediate filament proteins. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262:1401–1407

Hartmann MD, Ridderbusch O, Zeth K, Albrecht R, Testa O, Woolfson DN, Sauer G, Dunin-Horkawicz S, Lupas AN, Alvarez BH (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. Proc Natl Acad Sci U S A 106:16950–16955

Hartmann MD, Grin I, Dunin-Horkawicz S, Deiss S, Linke D, Lupas AN, Hernandez Alvarez B (2012) Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. Proc Natl Acad Sci U S A 109:20907–20912

Hartmann MD, Dunin-Horkawicz S, Hulko M, Martin J, Coles M, Lupas AN (2014) A soluble mutant of the transmembrane receptor Af1503 features strong changes in coiled-coil periodicity. J Struct Biol 186:357–366

Hartmann MD, Mendler CT, Bassler J, Karamichali I, Ridderbusch O, Lupas AN, Hernandez Alvarez B (2016) α/β coiled coils. eLife 5:e11861

Haslberger T, Weibezahn J, Zahn R, Lee S, Tsai FT, Bukau B, Mogk A (2007) M domains couple the ClpB threading motor with the DnaK chaperone activity. Mol Cell 25:247–260

Hazelbauer GL, Falke JJ, Parkinson JS (2008) Bacterial chemoreceptors: high-performance signaling in networked arrays. Trends Biochem Sci 33:9–19

Hernandez Alvarez B, Hartmann MD, Albrecht R, Lupas AN, Zeth K, Linke D (2008) A new expression system for protein crystallization using trimeric coiled-coil adaptors. Protein Eng Des Sel 21:11–18

Herrmann H, Bar H, Kreplak L, Strelkov SV, Aebi U (2007) Intermediate filaments: from cell architecture to nanomechanics. Nat Rev Mol Cell Biol 8:562–573

Hicks MR, Walshaw J, Woolfson DN (2002) Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. J Struct Biol 137:73–81

Hitchcock-DeGregori SE, Barua B (2017) Tropomyosin structure, function, and interactions: a dynamic regulator. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, Schultz A, Martin J, Schultz JE, Lupas AN, Coles M (2006) The HAMP domain structure implies helix rotation in transmembrane signaling. Cell 126:929–940

Kanchan K, Linder J, Winkler K, Hantke K, Schultz A, Schultz JE (2010) Transmembrane signaling in chimeras of the Escherichia coli aspartate and serine chemotaxis receptors and bacterial class III adenylyl cyclases. J Biol Chem 285:2090–2099

Khazina E, Weichenrieder O (2009) Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. Proc Natl Acad Sci U S A 106:731–736

Khazina E, Truffault V, Buttner R, Schmidt S, Coles M, Weichenrieder O (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. Nat Struct Mol Biol 18:1006–1014

Koiwai K, Hartmann MD, Linke D, Lupas AN, Hori K (2016) Structural basis for toughness and flexibility in the C-terminal passenger domain of an acinetobacter trimeric autotransporter adhesin. J Biol Chem 291:3705–3724

Korndorfer IP, Dommel MK, Skerra A (2004) Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. Nat Struct Mol Biol 11:1015–1020

Kreplak L, Bar H, Leterrier JF, Herrmann H, Aebi U (2005) Exploring the mechanical behavior of single intermediate filaments. J Mol Biol 354:569–577

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

Landschulz WH, Johnson PF, Mcknight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764

Lavigne P, Kondejewski LH, Houston MEJR, Sonnichsen FD, Lix B, Skyes BD, Hodges RS, Kay CM (1995) Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization. J Mol Biol 254:505–520

Lee S, Sowa ME, Watanabe YH, Sigler PB, Chiu W, Yoshida M, Tsai FTF (2003) The structure of clpB: a molecular chaperone that rescues proteins from an aggregated state. Cell 115:229–240

Lee S, Sowa ME, Choi JM, Tsai FTF (2004) The ClpB/Hsp104 molecular chaperone – a protein disaggregating machine. J Struct Biol 146:99–105

Leo JC, Lyskowski A, Hattula K, Hartmann MD, Schwarz H, Butcher SJ, Linke D, Lupas AN, Goldman A (2011) The structure of E. coli IgG-binding protein D suggests a general model for bending and binding in trimeric autotransporter adhesins. Structure 19:1021–1030

Leo JC, Grin I, Linke D (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos Trans R Soc Lond Ser B Biol Sci 367:1088–1101

Leroux MR, Fandrich M, Klunker D, Siegers K, Lupas AN, Brown JR, Schiebel E, Dobson CM, Hartl FU (1999) MtGimC, a novel archaeal chaperone related to the eukaryotic chaperonin cofactor GimC/prefoldin. EMBO J 18:6730–6743

Lesne E, Krammer EM, Dupre E, Locht C, Lensink MF, Antoine R, Jacob-Dubuisson F (2016) Balance between coiled-coil stability and dynamics regulates activity of BvgS sensor kinase in bordetella. MBio 7:e02089

Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA (2006) Trimeric autotransporter adhesins: variable structure, common function. Trends Microbiol 14:264–270

Liu J, Zheng Q, Deng Y, Cheng CS, Kallenbach NR, Lu M (2006) A seven-helix coiled coil. Proc Natl Acad Sci U S A 103:15457–15462

Lowe EC, Basle A, Czjzek M, Firbank SJ, Bolam DN (2012) A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. Proc Natl Acad Sci U S A 109:7298–7303

Lundin VF, Stirling PC, Gomez-Reino J, Mwenifumbo JC, Obst JM, Valpuesta JM, Leroux MR (2004) Molecular clamp mechanism of substrate binding by hydrophobic coiled-coil residues of the archaeal chaperone prefoldin. Proc Natl Acad Sci U S A 101:4367–4372

Lupas A (1996) Coiled coils: new structures and new functions. Trends Biochem Sci 21:375–382

Lupas AN, Gruber M (2005) The structure of alpha-helical coiled coils. Adv Protein Chem 70:37–78

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

Lupas AN, Bassler J, Dunin-Horkawicz S (2017) The structure and topology of α-helical coiled coils. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Malito E, Biancucci M, Faleri A, Ferlenghi I, Scarselli M, Maruggi G, Lo Surdo P, Veggi D, Liguori A, Santini L, Bertoldi I, Petracca R, Marchi S, Romagnoli G, Cartocci E, Vercellino I, Savino S, Spraggon G, Norais N, Pizza M, Rappuoli R, Masignani V, Bottomley MJ (2014) Structure of the meningococcal vaccine antigen NadA and epitope mapping of a bactericidal antibody. Proc Natl Acad Sci U S A 111:17128–17133

Martin J, Gruber M, Lupas AN (2004) Coiled coils meet the chaperone world. Trends Biochem Sci 29:455–458

Marx A, Hoenger A, Mandelkow E (2009) Structures of kinesin motor proteins. Cell Motil Cytoskeleton 66:958–966

Mason JM, Muller KM, Arndt KM (2007) Positive aspects of negative design: simultaneous selection of specificity and interaction stability. Biochemistry 46:4804–4814

Matamouros S, Hager KR, Miller SI (2015) HAMP domain rotation and tilting movements associated with signal transduction in the PhoQ sensor kinase. MBio 6:e00616–e00615

Mechaly AE, Sassoon N, Betton JM, Alzari PM (2014) Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. PLoS Biol 12:e1001776

Milburn MV, Prive GG, Milligan DL, Scott WG, Yeh J, Jancarik J, Koshland DEJR, Kim SH (1991) Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. Science 254:1342–1347

Miserez A, Guerette PA (2013) Phase transition-induced elasticity of alpha-helical bioelastomeric fibres and networks. Chem Soc Rev 42:1973–1995

Misura KM, Scheller RH, Weis WI (2000) Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex. Nature 404:355–362

Molnar KS, Bonomi M, Pellarin R, Clinthorne GD, Gonzalez G, Goldberg SD, Goulian M, Sali A, Degrado WF (2014) Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. Structure 22:1239–1251

Munson M, Chen X, Cocina AE, Schultz SM, Hughson FM (2000) Interactions within the yeast t-SNARE Sso1p that control SNARE complex assembly. Nat Struct Biol 7:894–902

Neiditch MB, Federle MJ, Pompeani AJ, Kelly RC, Swem DL, Jeffrey PD, Bassler BL, Hughson FM (2006) Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell 126:1095–1108

Ohtaki A, Kida H, Miyata Y, Ide N, Yonezawa A, Arakawa T, Iizuka R, Noguchi K, Kita A, Odaka M, Miki K, Yohda M (2008) Structure and molecular dynamics simulation of archaeal pre-

foldin: the molecular mechanism for binding and recognition of nonnative substrate proteins. J Mol Biol 376:1130–1141

Olia AS, Casjens S, Cingolani G (2009) Structural plasticity of the phage P22 tail needle gp26 probed with xenon gas. Protein Sci 18:537–548

Palecek JJ, Gruber S (2015) Kite proteins: a superfamily of SMC/Kleisin partners conserved across bacteria, archaea, and eukaryotes. Structure 23:2183–2190

Parkinson JS (2010) Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. Annu Rev Microbiol 64:101–122

Parry DAD (2005) Microdissection of the sequence and structure of intermediate filament chains. Adv Protein Chem 70:113–142

Parry DAD (2014) Fifty years of fibrous protein research: a personal retrospective. J Struct Biol 186:320–334

Parry DAD, Fraser RDB, Squire JM (2008) Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J Struct Biol 163:258–269

Qin W, Wang L, Lei L (2015) New findings on the function and potential applications of the trimeric autotransporter adhesin. Antonie Van Leeuwenhoek 108:1–14

Rackham OJ, Madera M, Armstrong CT, Vincent TL, Woolfson DN, Gough J (2010) The evolution and structure prediction of coiled coils across all genomes. J Mol Biol 403:480–493

Rao T, Ruiz-Gomez G, Hill TA, Hoang HN, Fairlie DP, Mason JM (2013) Truncated and helix-constrained peptides with high affinity and specificity for the cFos coiled-coil of AP-1. PLoS One 8:e59415

Saita E, Abriata LA, Tsai YT, Trajtenberg F, Lemmin T, Buschiazzo A, Dal Peraro M, De Mendoza D, Albanesi D (2015) A coiled coil switch mediates cold sensing by the thermosensory protein DesK. Mol Microbiol 98:258–271

Schultz JE, Natarajan J (2013) Regulated unfolding: a basic principle of intraprotein signaling in modular proteins. Trends Biochem Sci 38:538–545

Schultz JE, Kanchan K, Ziegler M (2015) Intraprotein signal transduction by HAMP domains: a balancing act. Int J Med Microbiol 305:243–251

Siegert R, Leroux MR, Scheufler C, Hartl FU, Moarefi I (2000) Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. Cell 103:621–632

Song HK, Hartmann C, Ramachandran R, Bochtler M, Behrendt R, Moroder L, Huber R (2000) Mutational studies on HslU and its docking mode with HslV. Proc Natl Acad Sci U S A 97:14103–14108

Sousa MC, Trame CB, Tsuruta H, Wilbanks SM, Reddy VS, Mckay DB (2000) Crystal and solution structures of an HsIUV protease-chaperone complex. Cell 103:633–643

Squire JM, Paul DM, Morris EP (2017) Myosin and actin filaments in muscle: structures and interactions. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Stetefeld J, Jenny M, Schulthess T, Landwehr R, Engel J, Kammerer RA (2000) Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. Nat Struct Biol 7:772–776

Strelkov SV, Burkhard P (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J Struct Biol 137:54–64

Sutton RB, Fasshauer D, Jahn R, Brunger AT (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. Nature 395:347–353

Swain KE, Falke JJ (2007) Structure of the conserved HAMP domain in an intact, membrane-bound chemoreceptor: a disulfide mapping study. Biochemistry 46:13684–13695

Swain KE, Gonzalez MA, Falke JJ (2009) Engineered socket study of signaling through a four-helix bundle: evidence for a yin-yang mechanism in the kinase control module of the aspartate receptor. Biochemistry 48:9266–9277

Szczesny P, Lupas A (2008) Domain annotation of trimeric autotransporter adhesins–daTAA. Bioinformatics 24:1251–1256

Szczesny P, Linke D, Ursinus A, Bar K, Schwarz H, Riess TM, Kempf VA, Lupas AN, Martin J, Zeth K (2008) Structure of the head of the Bartonella adhesin BadA. PLoS Pathog 4:e1000119

Tao Y, Strelkov SV, Mesyanzhinov VV, Rossmann MG (1997) Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. Structure 5:789–798

Tebo AG, Pecoraro VL (2015) Artificial metalloenzymes derived from three-helix bundles. Curr Opin Chem Biol 25:65–70

Trybus KM, Freyzon Y, Faust LZ, Sweeney HL (1997) Spare the rod, spoil the regulation: necessity for a myosin rod. Proc Natl Acad Sci U S A 94:48–52

Tskhovrebova L, Trinick J (2017) Titin and nebulin in thick and thin filament length regulation. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Utsumi R, Brissette RE, Rampersaud A, Forst SA, Oosawa K, Inouye M (1989) Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate. Science 245:1246–1249

Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, Cowan NJ (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell 93:863–873

Verver DE, Hwang GH, Jordan PW, Hamer G (2016) Resolving complex chromosome structures during meiosis: versatile deployment of Smc5/6. Chromosoma 125:15–27

Walton TA, Sousa MC (2004) Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. Mol Cell 15:367–374

Wang C, Sang J, Wang J, Su M, Downey JS, Wu Q, Wang S, Cai Y, Xu X, Wu J, Senadheera DB, Cvitkovitch DG, Chen L, Goodman SD, Han A (2013) Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains. PLoS Biol 11:e1001493

Watts KJ, Johnson MS, Taylor BL (2011) Different conformations of the kinase-on and kinase-off signaling states in the Aer HAMP domain. J Bacteriol 193:4095–4103

Wood CW, Bruning M, Ibarra AA, Bartlett GJ, Thomson AR, Sessions RB, Brady RL, Woolfson DN (2014) CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinformatics 30:3029–3035

Woolfson DN (2005) The design of coiled-coil structures and assemblies. Adv Protein Chem 70:79–112

Woolfson DN, Bartlett GJ, Bruning M, Thomson AR (2012) New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. Curr Opin Struct Biol 22:432–441

Woolfson DN (2017) Coiled-coil design: updated and upgraded. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Xiao L, Zhou L, Sun C, Feng X, Du C, Gao Y, Ji Q, Yang S, Wang Y, Han W, Langford PR, Lei L (2012) Apa is a trimeric autotransporter adhesin of Actinobacillus pleuropneumoniae responsible for autoagglutination and host cell adherence. J Basic Microbiol 52:598–607

Yeo HJ, Cotter SE, Laarmann S, Juehne T, St Geme JW IIIrd, Waksman G (2004) Structural basis for host recognition by the Haemophilus influenzae Hia autotransporter. EMBO J 23:1245–1256

Zhang F, Wu Z, Zhang P, Tian G, Finley D, Shi Y (2009) Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from Methanocaldococcus jannaschii. Mol Cell 34:485–496

Zhou Q, Ames P, Parkinson JS (2011) Biphasic control logic of HAMP domain signalling in the Escherichia coli serine chemoreceptor. Mol Microbiol 80:596–611

Zhou Q, Lai Y, Bacaj T, Zhao M, Lyubimov AY, Uervirojnangkoorn M, Zeldin OB, Brewster AS, Sauter NK, Cohen AE, Soltis SM, Alonso-Mori R, Chollet M, Lemke HT, Pfuetzner RA, Choi UB, Weis WI, Diao J, Sudhof TC, Brunger AT (2015) Architecture of the synaptotagmin-SNARE machinery for neuronal exocytosis. Nature 525:62–67

Zhu Y, Inouye M (2003) Analysis of the role of the EnvZ linker region in signal transduction using a chimeric Tar/EnvZ receptor protein, Tez1. J Biol Chem 278:22812–22819

Zhu L, Bolhuis PG, Vreede J (2013) The HAMP signal relay domain adopts multiple conformational states through collective piston and tilt motions. PLoS Comput Biol 9:e1002913

Chapter 4
The Structure and Topology of α-Helical Coiled Coils

Andrei N. Lupas, Jens Bassler, and Stanislaw Dunin-Horkawicz

Contents

Abstract α-Helical coiled coils constitute one of the most diverse folds yet described. They range in length over two orders of magnitude; they form rods, segmented ropes, barrels, funnels, sheets, spirals, and rings, which encompass anywhere from two to more than 20 helices in parallel or antiparallel orientation; they assume different helix crossing angles, degrees of supercoiling, and packing geometries. This structural diversity supports a wide range of biological functions, allowing them to form mechanically rigid structures, provide levers for molecular motors, project domains across large distances, mediate oligomerization, transduce conformational changes and facilitate the transport of other molecules. Unlike almost any other protein fold known to us, their structure can be computed from parametric

A.N. Lupas (✉) • J. Bassler
Department of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
e-mail: andrei.lupas@tuebingen.mpg.de; jens.bassler@tuebingen.mpg.de

S. Dunin-Horkawicz
Structural Bioinformatics Laboratory, Centre of New Technologies, University of Warsaw, 02-097 Warsaw, Poland
e-mail: s.dunin-horkawicz@cent.uw.edu.pl

equations, making them an ideal model system for rational protein design. Here we outline the principles by which coiled coils are structured, review the determinants of their folding and stability, and present an overview of their diverse architectures.

4.1 Introduction

Coiled-coil proteins were the earliest biological macromolecules studied at high resolution. Following the discovery of X-rays at the end of the nineteenth century, it was soon recognized that these provided a powerful new tool for probing the structure of biological matter. In an X-ray beam, crystalline materials produce diffraction patterns from which the position of individual atoms can be inferred, an insight that ushered in the era of structural biology at atomic resolution. Proteins remained too large for this method for several decades, requiring many technological and computational developments before the first atomic structure could be determined (myoglobin; Kendrew et al. 1958), but protein fibres were rapidly established as a rewarding target for X-ray diffraction. From the mid 1920s on, William Astbury probed systematically many natural fibres, including wool in the native and denatured state, porcupine quills, horn, nails, muscle, tendons, even DNA (Astbury 1938, 1946). He discovered that the diffraction spectra of protein fibres fell into a small number of spectral forms: an α-form exemplified by wool, a β-form exemplified by silk, and third form corresponding to tendon. He further found that proteins of the α-form could be converted to the β-form by stretching, or to yet another form, cross-β, by heat denaturation ("supercontraction"). All these forms showed simple diffraction patterns, illustrating that the underlying fibres were built by the regular repetition of simple structural motifs, a realization that invited numerous modelling efforts.

The most common diffraction pattern was the α-form, generated by a group of proteins that Astbury came to refer to as 'k-m-e-f', for keratin, myosin, epidermin, and fibrinogen. The hallmark of this pattern were prominent meridional arcs at 5.15 Å, indicating the repeating unit of the structure. Numerous attempts were made to derive molecular models from this pattern, including efforts by Astbury himself. These led to a race for the correct structural model between Bragg, Perutz and Kendrew at Cambridge and Pauling and Corey at Caltech (Judson 1979), a race that Linus Pauling won in 1950 with the announcement of the α-helix (Pauling and Corey 1950; Pauling et al. 1951). His structure, however, did not show a periodicity of 5.15 Å, as in the fibre diffraction pattern of the k-m-e-f proteins, but rather of 5.4 Å. Pauling had taken the confidence to ignore the meridional arcs of the α-form because the diffraction pattern of a synthetic fibre, poly-γ-methyl-L-glutamate, which was clearly of the α-type, also showed reflections at 5.4 Å away from the meridian. Pauling and Corey's α-helix was rapidly confirmed by diffraction data from the Cambridge group, but the meridional arcs of k-m-e-f proteins remained

unexplained and continued to preoccupy both groups. Undoubtedly these proteins were α-helical, but their constituent helices had to be distorted in some way relative to the ideal structure. Independently of each other, Linus Pauling at Caltech and Francis Crick at Cambridge found a solution to this problem through packing-induced superhelical distortions of the α-helices in the native proteins (Crick 1952, 1953a, b; Pauling and Corey 1953). Pauling called these supercoiled structures compound helices, Crick called them coiled coils.

In modelling the α-helix, Pauling and Corey (1950) had understood that the structure did not require an integer or a ratio of small integers for the number of residues per turn. Nevertheless, they had allowed for the possibility that, in a crystalline arrangement, the regular packing of the helices could favour deformations into configurations with a rational number of residues per turn, such as 11/3 (11 residues over 3 turns), 15/4, and 18/5. Two years later they returned to this idea, proposing that such superhelical deformation, resulting from the α-helices twisting around each other, could account for the meridional arc in the diffraction spectrum of k-m-e-f proteins (Pauling and Corey 1953). They discussed several sequence periodicities (4/1, 7/2, 15/4) with supercoil twists in the same or in the opposite sense to those of the constituent helices. They also discussed a number of different stoichiometries, including one with six helices coiling around a straight seventh one. This article did not offer quantitative parameters for the model structures and did not consider sidechain packing, proposing instead that supercoiling resulted from the exact repetition of short sequences that caused periodic fluctuations in backbone hydrogen-bond lengths.

Whereas Pauling and Corey only considered backbone configurations in their article, Crick placed sidechain geometry at the centre of his modelling effort (Crick 1952, 1953a, b). His key insight was that α-helices, when twisted around each other at an angle of about 20°, would interlock their sidechains systematically along the core of the structure, repeating the same interactions every seven residues (or two turns of the α-helix). Crick offered a full parameterization for such a structure with sequence periodicity of 7/2, which he called the 'coiled-coil', and referred to the systematic interlocking of sidechains as 'knobs' into 'holes'. He conjectured that the energy required to distort the helices could be provided by the packing interactions of the sidechains, if these were hydrophobic, leading him to predict that the sequence of coiled coils would show a periodic recurrence of hydrophobic residues with a periodicity of $7/2 = 3.5$ (a periodicity that in time became known as the 'heptad repeat').

Although the heptad periodicity predicted by Crick was only confirmed more than 20 years later with the determination of the tropomyosin sequence (Stone et al. 1975) and knobs-into-holes packing with the structure of influenza hemagglutinin (Wilson et al. 1981), the elegance and simplicity of Crick's parameterization was immediately convincing and became the canonical account of coiled-coil structure. As an intriguing aside, the first coiled-coil protein of known sequence was murein lipoprotein (Braun and Bosch 1972) and the first with a high-resolution structure catabolite gene activator protein (McKay and Steitz 1981), but the coiled coils in these proteins were only recognized years later.

Following Crick's model of the coiled coil, the number of papers on this topic saw an explosive increase, also relative to other topics in the rapidly expanding field of protein biochemistry, but started losing ground after the early 1960s (Fig. 4.1). Interest was rekindled by the growing availability of protein sequences, which opened coiled coils to computational sequence analysis and modelling (e.g. McLachlan and Stewart 1975, 1976; Parry 1975; Parry et al. 1977), and by the first high-resolution structures, which led Carolyn Cohen and David Parry to propose that an understanding of coiled coils would not only be relevant for long fibrous proteins, but also for globular and membrane proteins (Cohen and Parry 1986, 1990). Their insight was proven visionary just 2 years later with the leucine zipper hypothesis (Landschulz et al. 1988), which fundamentally altered the perception of coiled coils in protein science (see e.g. Cohen and Parry 1994). This led to a second phase of explosive growth for this topic (Fig. 4.1) and coiled coils and leucine zippers are now household terms in protein science. There have been many excellent reviews on their various aspects over the years (Lupas 1996; Cohen 1998; Burkhard et al. 2001; Gruber and Lupas 2003; Mason and Arndt 2004; Lupas and Gruber 2005; Woolfson 2005; Parry et al. 2008), so we will place more emphasis here on developments of the last decade and refer our readers to these reviews for further information.
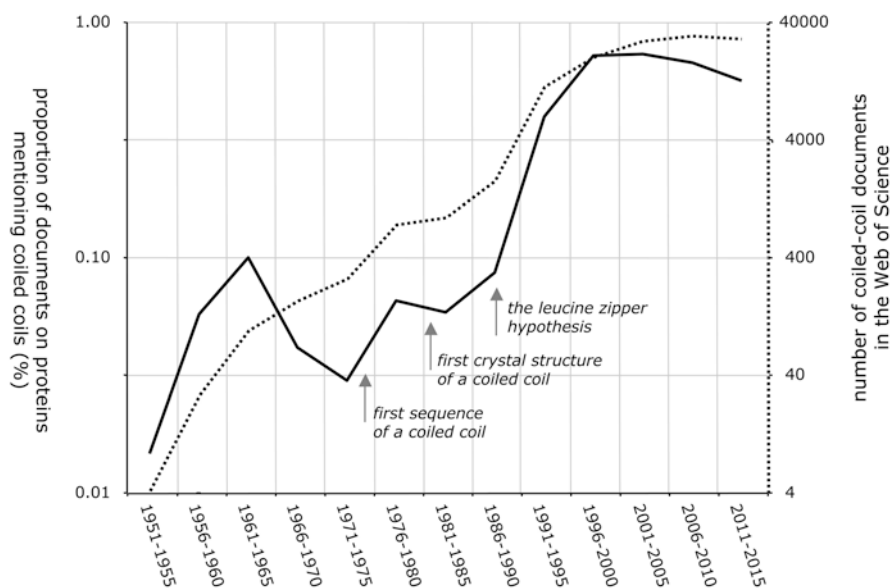
Fig. 4.1 Number of coiled-coil publications over the years. The number of documents in the Web of Science (Thomson Reuters) obtained with the Topic query [("coiled coil*" OR "leucine zipper") AND protein*] over 5-year intervals starting with 1951–1955 are shown with a *dashed line* (*right axis*). These numbers, divided by the number of documents obtained with the Topic query [protein*] over the same intervals, are shown with a *solid line* (*left axis*). More than 5000 coiled-coil publications have appeared every year for the last two decades, amounting to about 0.5 % of publications in the protein sciences

4.2   Structural Parameters

4.2.1   The Standard Model

Coiled coils are α-helical structures in which helices are wound around each other to form superhelical bundles. They usually consist of two or three helices in parallel or antiparallel orientation, but structures with seven and more helices have been determined. They are usually oligomers either of the same (homo) or of different chains (hetero), but can also be formed by consecutive helices from the same polypeptide chain, which in that case almost always have an antiparallel orientation.

The constituent helices of coiled coils interact via a knobs-into-holes geometry of amino-acid sidechains at their interface (Crick 1952, 1953b). In this geometry, a residue from one helix (knob) packs into a space surrounded by four sidechains of the facing helix (hole). Because the residue thus comes to be located next to the equivalent residue from the facing helix, this geometry is sometimes referred to as in-register. This geometry contrasts with the more irregular packing of helices in globular proteins and non-coiled-coil bundles. In that packing mode, referred to as ridges-into-grooves (Chothia et al. 1977, 1981), a residue packs above or beneath the equivalent residue from the facing helix and is therefore also called out-of-register. The two packing modes impose fundamentally different constraints on the α-helices. Whereas ridges-into-grooves packing can be formed with undistorted helices and a multitude of sidechain positions, the regular meshing of knobs-into-holes packing requires precisely recurrent positions of the side-chains every seven residues along the helix interface (the heptad repeat). Because these seven positions, arranged over two turns of the helix, yield a periodicity of 3.5 residues per turn and an undistorted helix has 3.63 residues per turn, the helices must bend to reduce the periodicity to 3.5 relative to a central axis. This bending is called supercoiling and occurs in the sense opposite to that of the individual helices; this means that, because α-helices are right-handed, the supercoil of heptad coiled coils must be left-handed. The regularity of side-chain positions is reflected at the level of their biophysical properties: coiled coils show seven-residue sequence repeats whose positions are labelled *a–g*; the core-forming positions (*a* and *d*) are usually occupied by hydrophobic residues, whereas the remaining, solvent-exposed positions (*b*, *c*, *e*, *f*, and *g*) are dominated by hydrophilic residues.

The consequence of the regular nature of coiled coils is that their structures can be described fully by parametric equations (Crick 1953a; Offer et al. 2002; Lupas and Gruber 2005). The coiled-coil parameters characterize individual helices and their orientation in a superhelical bundle (Fig. 4.2). These parameters can be used to calculate a Cα trace of a coiled-coil bundle, which upon backbone reconstruction and side-chain placement, can serve as a starting point for design and modelling procedures (see Chap. 2 by Woolfson 2017, in this issue).
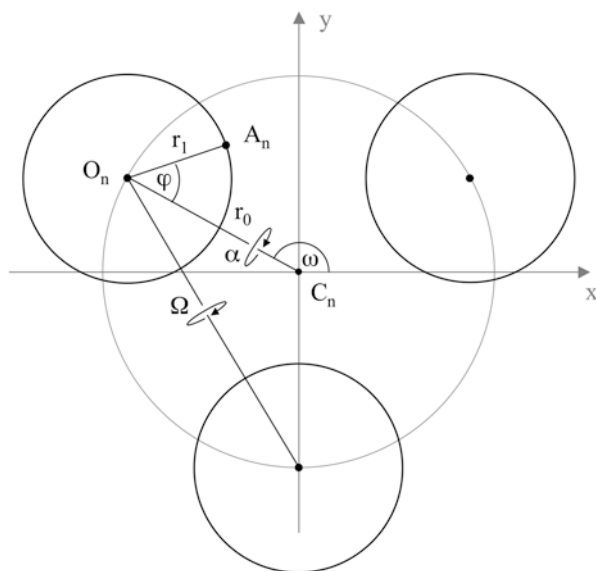
Fig. 4.2 Schematic representation of a trimeric coiled coil, showing the main parameters. O_n marks the centre of one α-helix, A_n the Cα position of a constituent residue, and C_n the superhelix axis. The distance required for the superhelix to complete a full turn is called the pitch, and the angle of a helix relative to the superhelical axis is α, the pitch angle (sometimes also called superhelix crossing angle or tilt angle). The angle between two neighbouring helices is Ω, the pairwise helix-crossing angle. The vector connecting the centre of a helix to the superhelical axis gives r_0, the superhelix radius, and that connecting the centre of a helix to the Cα carbons of its constituent residues gives r_1, the α-helix radius. The angle between the α-helix radius and superhelix radius vectors for the same residue is φ, the positional orientation angle, or Crick angle (which is sometimes confusingly denoted α as well); it gives the location of a given residue relative to the supercoil axis. The angle between the α-helix radius vectors for two consecutive residues is the phase shift of the α-helix ($\Delta\varphi$) and the angle between two consecutive superhelix radius vectors is the phase shift of the supercoil ($\Delta\omega$)

4.2.2   Prediction and Analysis Programs

Thanks to the regularity of the repeating pattern, coiled coils can be predicted directly from their sequence to a level of detail that permits the assignment of individual residues to the positions of the heptad repeat. The most popular tools for such predictions follow the idea of scoring the sequence against a matrix of residue frequencies derived from known coiled coils (Parry 1982). This approach was implemented in the **COILS** program by substituting residue preferences for frequencies (i.e. dividing residue frequencies by their background frequencies in the sequence database), introducing a scanning window, and scaling scores against reference databases to obtain probabilities (Lupas et al. 1991; Gruber et al. 2006). A variant

of this approach, using pairwise residue correlations, was implemented in the program **PairCoil** (Berger et al. 1995; McDonnell et al. 2006). More recently, another approach based on Hidden Markov Models led to the development of **MARCOIL** (Delorenzi and Speed 2002), which operates without a scanning window and appears to offer the best combination of sensitivity and speed among the currently available methods (Gruber et al. 2006). Conceptually similar methods have also been implemented to predict the oligomerization state of coiled coils from sequence data, yielding the programs **SCORER** (Woolfson and Alber 1995; Armstrong et al. 2011) and **Multicoil** (Wolf et al. 1997; Trigg et al. 2011), and, through use of support vector machine-based classification, **PrOCoil** (Mahrenholz et al. 2011). These programs all distinguish only between parallel dimers and trimers, and are unable to identify coiled coils that are neither, thus being somewhat limited in their use. A step forward has been taken with **LOGICOIL** (Vincent et al. 2013), a program that combines Bayesian variable selection with multinomial probit regression for prediction; this program is able to additionally distinguish antiparallel dimers and tetramers, thus considerably extending its range. For all programs discussed here, it is important to keep in mind that they are designed to analyze unbroken heptad repeats and therefore have limited applicability to coiled coils with other periodicities (see Sect. 4.2.4). In addition, the sequences of coiled coils are similar to those of natively unstructured proteins, leading to an overlap between the two types of predictions. The reasons for this similarity and ways to deal with the resulting false negative and false positive predictions will be discussed in Sect. 4.3.2.

The ability to describe coiled coils with parametric equations has also led to the development of a large number of useful programs that analyse and model coiled-coil structures. **Socket** (Walshaw and Woolfson 2001) detects knobs-into-holes packing and **Twister** (Strelkov and Burkhard 2002), **SamCC** (Dunin-Horkawicz and Lupas 2010a), and **CCCP fit** (Grigoryan and DeGrado 2011) quantify the properties of experimental structures, while **BeammotifCC** (Offer et al. 2002), **CCCP generate** (Grigoryan and DeGrado 2011), and **CCBuilder** (Wood et al. 2014) allow users to generate models for coiled coils with predefined properties. **Socket** relies on a user-adjustable geometric definition of knobs-into-holes packing and on a simplified structure representation, considering sidechains only by their centre of mass. A scan of the Protein Structure Databank with **Socket** allowed Woolfson and coworkers to classify numerous coiled-coil structures according to their architecture in a relational database, **CC+** (Testa et al. 2009). In contrast to **Socket**, **Twister** ignores the sidechains completely and considers only Cα carbons of a coiled-coil structure to determine the position of the central axis for each helix which, in turn, determines the location of the supercoil axis. Once the axes are traced, all structural parameters defined by the Crick parameterization can be calculated at the resolution of individual residues. **Twister** is very well suited to track local fluctuations in coiled-coil structures, but is limited by the requirement that the helices be in a parallel orientation and symmetric (because only averaged results are emitted). This issue has been addressed in the **SamCC** program, which can be used to measure antiparallel and asymmetric coiled coils with four or more helices. A different

approach is used in the **CCCP** suite of programs, which does not divide the structure into consecutive layers, but rather globally fits Crick parameters to a given backbone structure. **CCCP** can not only be used to analyse structures, but also to produce coordinates for the main chains of coiled coils based on specified Crick parameters. Such modelling can be also performed with **BeammotifCC**, which is uniquely able to model transitions between regions of different periodicity. Finally, modelling of coiled coils can also be performed in **CCBuilder**, which is available as a web service only and provides the user with a complete pipeline for coiled coil design and model validation.

4.2.3   Coiled Coils with Variant Core Geometry

While most coiled coils follow the canonical knobs-into-holes packing described by Crick, two variant packing modes are sometimes observed in antiparallel structures. The first, complementary *x–da*, is brought about by global axial rotation of all helices, in the ideal case by about 26°. This rotation shifts the relative position of residues and leads to a variant of knobs-into-holes packing, in which the hydrophobic core is formed by three positions, rather than two. When viewed from the N-terminus of a helix, clockwise rotation moves position *d* into the centre of the core, position *a* outward and position *g* inward to yield an *a–d–g* core; counter-clockwise rotation has the opposite effect, moving position *a* to the centre and yielding an *a–d–e* core (Fig. 4.3a). In both cases, the positions of the extended hydrophobic core assume two distinct geometries: we refer to one geometry as *x*, where the sidechains point towards the centre of the coiled-coil bundle, and to the other as *da*, where two sidechains point side-ways, enclosing a central cavity. Importantly, *x* and *da* represent structural nomenclature and *da* should not be confused with positions *d* and *a* of a heptad repeat. In the *a–d–g* core, positions *g* and *a* assume *da* geometry, whereas position *d* assumes *x* geometry. Similarly, in the *a–d–e* core, positions *d* and *e* assume *da* geometry and position *a* assumes *x* geometry (Fig. 4.3a). Complementary *x–da* packing is almost invariably found in antiparallel four-helical bundles, because *x* and *da* alternate along the helices and therefore the geometrically most favourable packing is achieved when the two pairs of diagonally opposite helices run in inverse directions and the *x* positions in one pair are combined with *da* positions in the other. The only case known to us where complementary *x–da* packing occurs in a parallel coiled coil is in the HAMP domain, a homodimer of two consecutive helices in which complementarity is achieved by the N-terminal helix having an *a–d–e* core and the C-terminal one an *a–d–g* core (Hulko et al. 2006). Intriguingly, HAMP can assume both canonical and complementary *x–da* packing (Ferris et al. 2011), suggesting that it could relay conformational signals in transmembrane receptors by the concerted axial rotation of its helices (Mondejar et al. 2012; Ferris et al. 2012, 2014).
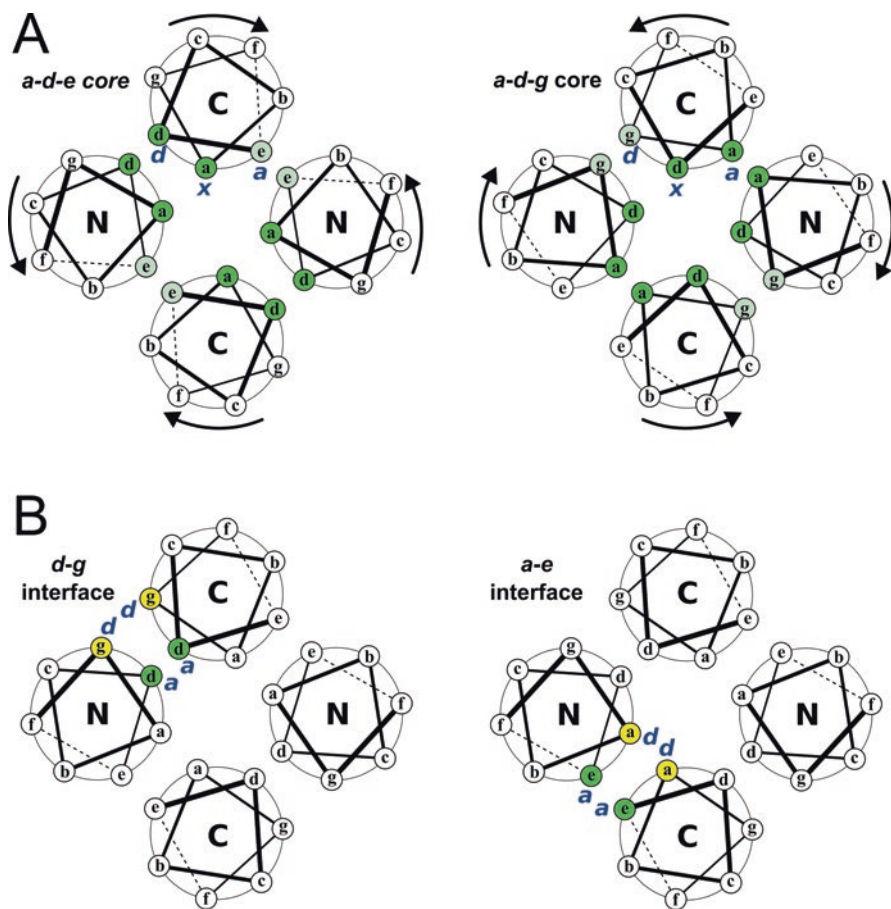
Fig. 4.3 Variant core geometries. (**a**) Complementary *x–da* packing via axial helix rotation in antiparallel four-helical coiled coils, showing helical wheel-diagrams of *a–d–e* (*left*) and *a–d–g* (right) hydrophobic cores. Canonical hydrophobic core positions *(a and d)* and positions co-opted to the core in *x–da* packing (*e* or *g*) are *highlighted*. (**b**) Alacoil interactions in the context of four-helical bundles. The heptad positions of the helices are denoted relative to the overall register of the bundle. The Alacoil interactions can occur along either the *d–g* (*left*) or the *a–e* (*right*) edge; the participating positions are *highlighted*

The extent to which helices are axially rotated is captured by the Crick angle of the standard model, which describes the orientation of a residue relative to the bundle axis. Measurements of Crick angles in experimentally determined structures using **SamCC** showed that coiled coils rarely assume an ideal *x–da* packing, in which helices are rotated by +26° or −26°. Rather, a continuum of axial rotational states was observed, predominantly in coiled coils with *a–d–e* cores (Dunin-Horkawicz and Lupas 2010a); in contrast, coiled coils with *a–d–g* cores are rarely observed and their rotation angles are typically <10° (Szczepaniak et al. 2014). A survey of natural and computed structures found that small hydrophobic residues in

position *e* and hydrophilic residues in position *g* favour *a–d–e* cores; the opposite distribution favours *a–d–g* cores (Dunin-Horkawicz and Lupas 2010b; Szczepaniak et al. 2014). Depending on the sidechain size of the residues co-opted into the core (*e* in the case of *a–d–e* cores, *g* for *a–d–g* cores) the cross-sections of the resulting bundles may range from square (e.g. GCN4-pV, Table 4.1) to distinctly rectangular (e.g. in Lac21 and GCN4-pAe$_{LV}$, where the co-opted positions *e* are occupied by alanines, Table 4.1).

The second variant packing mode, Alacoil, refers to an arrangement of two tightly associated, antiparallel α-helices (Gernert et al. 1995); we will address the unique case of an engineered, parallel Alacoil, which assembles into sheets, in Sect. 4.3.1. The exceptionally small distance of ~8 Å between helical axes is possible owing to the overrepresentation of small residues on one side of the hydrophobic core, typically alanine (hence the name). If the small residues occur in position *a* (defined with respect to the pairwise interaction), the resulting Alacoil is considered to be of the ferritin type, if in *d*, of the ROP type (Fig. 4.4; Table 4.1). The residues in the other core position (*d* in ferritins and *a* in ROPs) are typically large and interdigitate to form a continuous ridge. The two Alacoil types can also be distinguished by the relative shift between the interacting helices: when viewed from the side, helical turns in ROP-like structures appear to lie directly across each other, whereas in ferritin-like structures they appear offset (Gernert et al. 1995).

In the original description of the Alacoil (Gernert et al. 1995), all examples were pairwise helical interactions extracted from bundles comprising four α-helices. The heptad register used to describe them referred to the pairwise interaction, ignoring the register of the parent bundle. Since Alacoils can form along either the *a–e* or *d–g* edges of a bundle (Fig. 4.3b; Table 4.1), the pairwise register has to be related to the overall register on a case-by-case basis. This has proven a consistent source of confusion, further enhanced in many studies by a failure to distinguish between Alacoil interactions and *x–da* packing. Although some coiled coils with *x–da* packing resemble Alacoils in having a seam of small residues that leads to close, pairwise association of their helices, the two packing modes differ fundamentally in the axial rotation state of their helices with respect to the bundle. Whereas helices showing Alacoil interactions are essentially unrotated (Dunin-Horkawicz and Lupas 2010a), helices with *x–da* interactions are rotated by > ±10° with respect to the bundle axis (Szczepaniak et al. 2014). An example is provided by the four-helical bundle formed by the Lac repressor tetramer, Lac21, which was described as a ferritin-like Alacoil (Solan et al. 2002), but, in fact, shows complementary *x–da* packing with an *a–d–e* core, due to the ~16° counter-clockwise rotation of its helices (Table 4.1).

4.2.4   Non-heptad Coiled Coils

Coiled coils typically show a high degree of regularity, but are rarely without any discontinuities. The most common of these result from the insertion of three (stammer) or four (stutter) residues into the heptad repeat. Both are close enough to the

Table 4.1 Examples of coiled coils with Alacoil or *x–da* interactions

Name	PDB id	Architecture	Core geometry	av. axial rotation[a]	Min. pairwise interface distances (Å)[b]	Tight interface type	Comment
F-BAR	3Q0K	Hairpin	Alacoil	–	**B [153–177] – B [212–188]: 8.3**	ROP-like (*a–d*)	
pT26-6p	2WB7	Sheet of 4 helices	Type II bifaceted with Alacoil interfaces	–	**A [565–582] – A [604–587]: 7.7** **A [592–610]: A [638–620]: 7.6** **A [625–643] – A [666–648]: 7.4**	All three interfaces are ferritin-like	The two central helices are bifaceted (first *c–g/d–a*, second *d–a/e–b*) and cancel out each other's angular offset (see Sects. 4.3.1 and 4.4.2)
SMBP	3U8V	Four-helical bundle	Canonical with two Alacoil interfaces	0°	**A [6–20] – A [43–29]: 7.6** A [43–29] – A [51–65]: 9.2 **A [51–65] – A [88–74]: 7.6** A [88–74] – A [6–20]: 8.7	Ferritin-like (*d–g*)	Duplication of a hairpin that forms a His-zipper Alacoil
ROP	1ROP	Antiparallel homodimer of hairpins	Canonical with two Alacoil interfaces	0°	**A [28–7] – A [32–53]: 8.5** A [32–53] – B [53–32]: 9.8 **B [53–32] – B [7–28]: 8.5** B [7–28] – A [28–7]: 10.3	ROP-like (*a–e*)	

Table 4.1 (continued)

Name	PDB id	Architecture	Core geometry	av. axial rotation[a]	Min. pairwise interface distances (Å)[b]	Tight interface type	Comment
Ferritin	1FHA	Four-helical bundle	Irregular with one Alacoil interface	(too irregular)	A [16–41] – A [121–96]: 10.4 A [121–96] – A [129–154]: 9.0 A [129–154] – A [76–51]: 12.4 **A [76–51] – A [16–41]: 7.7**	Ferritin-like (a–e)	
GCN4-pAg	2B1F	Antiparallel homotetramer	Canonical with two Alacoil interfaces	–5°	**A [5–29] – B [29–5]: 7.9** B [29–5] – C [5–29]: 9.5 **C [5–29] – D [29–5]: 8.1** D [29–5] – A [5–29]: 9.4	Ferritin-like (d–g)	
Form I peptide	3J89	Parallel open-ended sheet	Type III bifaceted with both interfaces Alacoil		**A [11–29] – E [8–26]: 8.5**	ROP-like (g–d/e–b)	Only example of a parallel Alacoil
DSD	1G6U	Hexameric tube formed as a trimer of dimers (all subunits antiparallel)	Type III bifaceted with an Alacoil interface forming between dimers	–	interfaces of the central tube: **B/H/E [24–41] – G/D/A [41–24]: 7.7** G/D/A [41–24] – H/E/B [24–41]: 10.4	Ferritin-like (d–g)	Designed hairpin with one short and one long helix; forms an antiparallel three-helical dimer; three dimers assemble via their long helices into a six-helical tube buttressed by the short helices

LPD-12	3CAY	Antiparallel homodimer	Alacoil	–	**A [4–20] – B [20–4]: 7.9**	Intermediate between ferritin-like and ROP-like (both interface positions are alanine)	Synthetic peptide comprised of two nearly identical hendecads, yet the dimer forms with heptad periodicity (Alacoil geometry may only be compatible with heptads)
Lac21	1LBI	Antiparallel homotetramer	Complementary x–da (a–d–e core) = type I bifaceted	–16°	**A [356–340] – C [339–355]: 7.3**; C [339–355] – D [356–340]: 10.2; **D [356–340] – B [339–355]: 7.2**; B [339–355] – A [356–340]: 10.5	a–e	Square cross-section
GCN4-pAe_LV	2R2V	Antiparallel homotetramer	Complementary x–da (a–d–e core) = type I bifaceted	–8°	A [6–31] – B [31–6]: 11.2; **B [31–6] – C [6–31]: 7.9**; C [6–31] – D [31–6]: 11.2; **D [31–6] – A [6–31]: 7.8**	a–e	Square cross-section
GCN4-pVg	2B22	Antiparallel homotetramer	Complementary x–da (a–d–e core) = type I bifaceted	–21°	A/C [5–29] – B/D [29–5]: 10.1; B/D [29–5] – C/A [5–29]: 9.9	–	Square cross-section

(continued)

Table 4.1 (continued)

Name	PDB id	Architecture	Core geometry	av. axial rotation[a]	Min. pairwise interface distances (Å)[b]	Tight interface type	Comment
SARS	1ZV7	Antiparallel homotetramer	Complementary x–da (a–d–e core) = type I bifaceted	−16°	B [5–36] – A [36–5]: 8.9 / A [36–5] – C [5–36]: 10.1 / C [5–36] – D [36–5]: 8.9 / D [36–5] – B [5–36]: 9.4	–	Rectangular cross-section
HAMP	2L7H	Dimer of helix-loop-helix subunits (all helices parallel)	Complementary x–da (helices A1 and B1: a–d–e; helices A2 and B2: a–d–g) = type I bifaceted	helices A1 & B1: −20° / helices A2 & B2: 20°	A [283–296] – A [311–324]: 10.6 / A [311–324] – B [283–296]: 8.9 / B [283–296] – B [311–324]: 10.7 / B [311–324] – A [283–296]: 9.0	–	Rectangular cross-section; only example of a parallel bundle with complementary x–da packing

[a]Four-helical bundles only

[b]Letters denote chains in the structure from column B, numbers denote the respective residue range; tight interfaces with a minimal distance of 8.5 Å or less are shown *bold*

Fig. 4.4 Alacoil interactions between antiparallel helices. In ROP-type Alacoils (*left column*) small residues are localized in position *d* of the heptad repeat, whereas in Ferritin-type (*right column*) in position *a*. Positions occupied by small resides are *highlighted*. Note the difference in the relative axial shift between the interacting helices
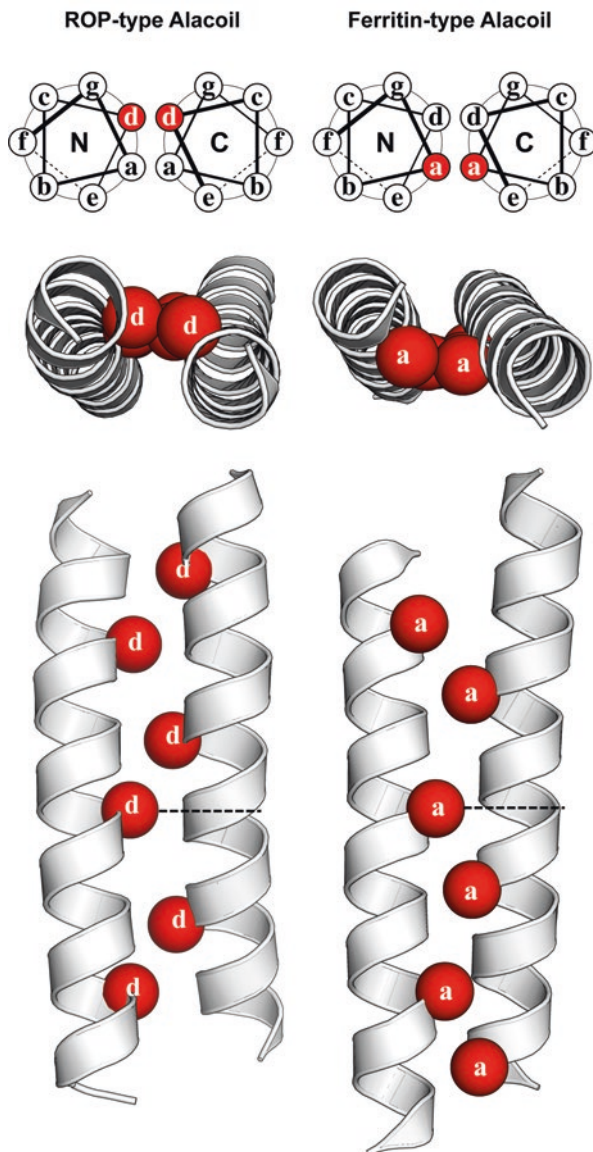
periodicity of canonical coiled coils (7/2 = 3.5) to be accommodated without disrupting the constituent helices. By inserting more than 3.5 residues, stutters raise the periodicity locally to 3.67 residues per turn (7 + 4 = 11 -> 11/3), leading to an unwinding of the left-handed supercoil; stammers have the opposite effect and reduce the periodicity to 3.33 (7 + 3 = 10 -> 10/3), causing overwinding. Stutters, in particular, are frequently seen in coiled coils and the 11-residue segments they generate are often referred to as hendecads. These local changes in supercoiling

affect core packing in the same way as the global rotation of the helices (Sect. 4.2.3), leading to the local adoption of *x–da* geometry: stutters are analogous to counter-clockwise rotation and stammers to clockwise rotation, as seen from the N-terminus of the helix. In coiled coils that are parallel and in-register, residues in positions *x* point towards the supercoil axis and residues in *da* interact in a ring around a central cavity; both depart from knobs-into-holes geometry to form knobs-to-knobs interactions. In dimeric coiled coils, these knobs-to-knobs interactions are particularly constrained, as residues in *x* point directly at each other, causing steric clashes if the residues are larger than glycine or alanine. Several strategies to acco-modate *x* layers are observed (Lupas and Gruber 2005): the coiled coils may become locally asymmetric to avoid clashes; they may form higher oligomers, where the centre of the bundle offers increasingly more space; they may move locally out of register towards ridges-into-grooves geometry; or they may form antiparallel bun-dles, which have more space for residues in *x* positions because sidechains are angled towards the N-terminus and thus point in opposite directions (also, here *x* and *da* geometries can be combined to form complementary *x–da* layers, as dis-cussed in Sect. 4.2.3).

Insertions may be delocalized over several heptads; for example, a stutter can result in higher periodicities like 18 residues over 5 turns (14 + 4 = 18 -> 18/5 = 3.6), or 25 residues over 7 turns (21 + 4 = 25 -> 25/7 = 3.57). Delocalization over more than one heptad also allows for accommodation of discontinuities other than stutters or stammers, such as insertions of 1 (skip) or 5 residues. In the context of a single heptad, these would lead to periodicities of 2.66 ((7 + 1)/3), 3.0 ((7 + 5)/4), or 4.0 ((7 + 1)/2; (7 + 5)/3), none of which fall into the range accessible to coiled coils. However, if delocalized over at least two heptads they produce periodicities such as 3.67 ((21 + 1)/6), 3.71 ((21 + 5)/7), 3.75 ((14 + 1)/4), or 3.8 ((14 + 5)/5), all of which can result in stable structures. Fundamentally, all periodicities accessible to coiled coils without disruption of the constituent helices can be described as combinations of three- and four-residue segments that start with a core residue (Hicks et al. 2002; Gruber and Lupas 2003). This is evidently true for heptads (7), stutters (4), and stammers (3), but also for skips, which in the context of a heptad are equivalent to two stutters (7 + 1 = 4 + 4), and 5-residue insertions, which, in the same context, are equivalent to three stutters (7 + 5 = 4 + 4 + 4). When elements of 3 and 4 alternate regularly, perfect knobs-into-holes packing is achieved. In places where consecutive elements of the same kind (3–3 or 4–4) occur, the packing locally requires knobs-to-knobs interactions.

There are limits to the periodicities that coiled coils can assume, resulting from the supercoil strain imposed on the constituent helices. The further a periodicity is from the 3.63 residues per turn of an undistorted helix, the stronger the supercoiling needed to form a coiled coil. At some point, the strain is sufficient to break the helix. Indeed, two crystal structures of stammers in heptad coiled coils show that the strain these introduce is accommodated by the formation of local 3_{10}-helical segments (Hartmann et al. 2009, 2016). It thus appears that 3.33 residues per turn (or 0.3 less than an undistorted helix) mark the lower boundary of periodicities accessible to coiled coils. By extension, we anticipate the upper boundary to be around 3.95.

Indeed, if a skip residue cannot be accommodated by delocalization over at least two heptads, the local periodicity of 4.0 it causes is too high for a coiled coil and the residue is looped out into a π-turn, leaving the remaining coiled coil largely unperturbed (Lupas 1996; Lupas and Gruber 2005).

If the insertion of 3 residues marks a lower limit for the supercoiling accessible to a coiled coil, what happens if 2x3 or 3x3 residues are inserted (corresponding to the deletion of 1 or insertion of 2 residues in a heptad frame, respectively)? Surprisingly, in trimeric coiled coils, these were recently shown to cause the local formation of short β-strands, which move the path of each chain in the trimer by 120° around the supercoil axis (Hartmann et al. 2016), resulting in an α/β coiled coil that represents a substantially new kind of fibre (Hartmann et al. 2016) (Fig. 4.5).

When the local discontinuities described here are repeated regularly along the helix, they lead to the formation of coiled coils whose periodicity deviates globally from the standard model. For these, the handedness of the supercoil is given by the difference of their periodicity to the 3.63 of an undistorted α-helix: coiled coils with periodicities of 3.4 (17/5) or 3.57 (25/7) are left-handed; with periodicities of 3.6 (18/5) or 3.67 (11/3) essentially straight, and with periodicities of 3.75 (15/4) or 3.8 (19/5) right-handed (Fig. 4.5). Since only heptads lead to continuous knobs-into-holes packing, all these periodicities cause the periodic formation of knobs-to-knobs interactions. Because of the constraints resulting from them, coiled coils that deviate globally from the heptad pattern are essentially never two-helical. Such coiled coils were anticipated by Pauling and Corey (1953), but not considered by Crick, who placed knobs-into-holes interactions at the centre of his model. Nevertheless, the Crick equations fully account for coiled coils with non-heptad periodicities and their implementation into computational tools like **CCCP** or **BeammotifCC**, can be used to measure and model such coiled coils. **BeammotifCC**, though, is the only tool that can model transitions between different types of periodicities such as seen in many natural coiled coils, for example in the rod of the trimeric adhesin YadA, which goes from 15/4 to 19/5 to 7/2 (Koretke et al. 2006).

4.3 Structural Determinants of Folding and Stability

4.3.1 Number and Orientation of Helices

The sidechains at the helical interfaces of coiled coils are the main determinants of their oligomeric state and of the orientation of the helices in the bundle. We will outline a number of factors here that can lead to a specific structural state, but individually these should be seen as context-dependent preferences, which may be overridden by other factors.

In parallel coiled coils, an important role is played by the core residues in positions *a* and *d*, which show distinct preferences for specific sidechains in two-, three-, and four-helical coiled coils. This is because in position *a* of dimeric coiled coils,
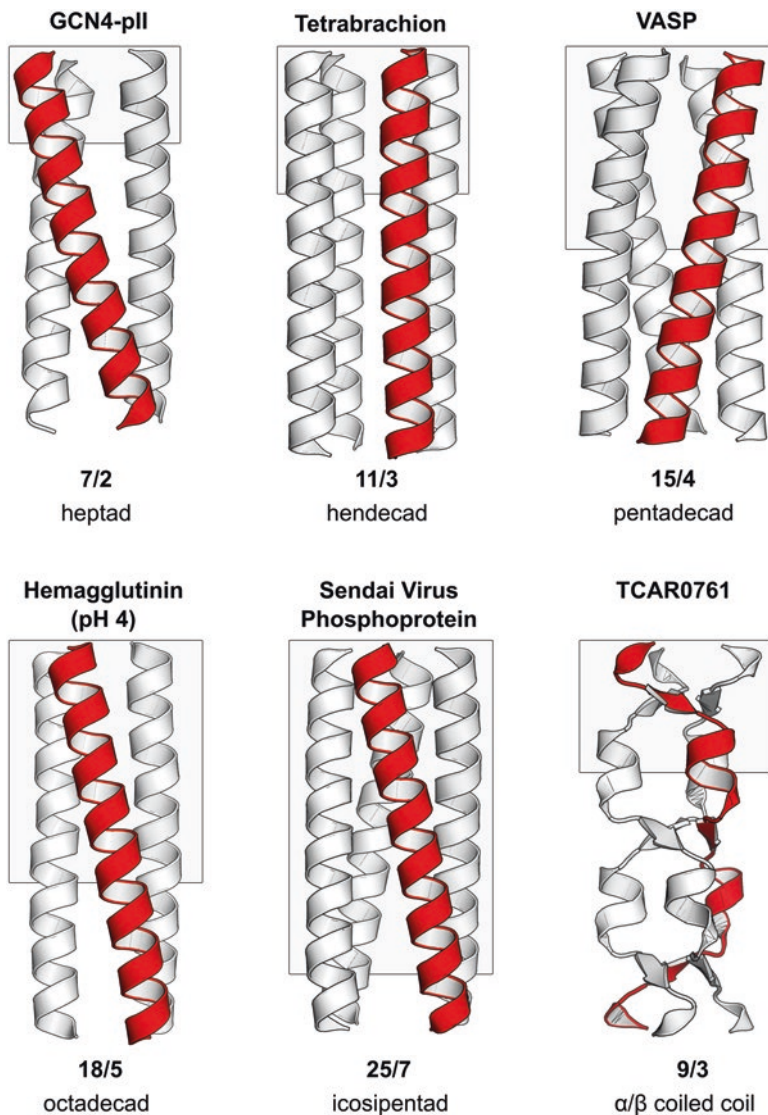
Fig. 4.5 Coiled coil structures with different periodicities. *Top row*: The left-handed supercoil of GCN4-pII (PDB: 1GCM) with a periodicity of seven residues over two helical turns (7/2), the straight helices of tetrabrachion (PDB: 1FE6) with 11/3 periodicity, and the right-handed supercoil of human VASP (PDB: 1USD) with 15/4 periodicity. *Bottom row*: Two mildly left-handed super-coils created by a 18/5 periodicity in the influenza hemagglutinin, pH 4 (PDB: 1HTM) and a 25/7 periodicity in the Sendai virus phosphoprotein (PDB: 1EZJ). In TCAR0761 from *Thermosinus carboxydivorans* (pdb: 5APZ), the 9/3 periodicity leads to the formation of an α/β coiled coil. *Grey* backgrounds indicate the extent of a single repeat

the Cα-Cβ vector of residues is parallel to that of the peptide bond facing them across the interface, while in position *d* it is perpendicular. This provides more space at the centre of the interface for position *a*, favouring the β-branched residues isoleucine and valine, and a more constrained central space for position *d*, favouring the γ-branched residue leucine. In trimeric coiled coils, the Cα-Cβ vectors of positions *a* and *d* have a similar, acute angle with respect to the interface and do not show a preference for specific sidechains. In tetrameric coiled coils, the geometry of the positions at the interface is reversed, with *a* having a perpendicular Cα-Cβ vector and a preference for leucine, and *d* having a parallel vector and a preference for isoleucine and valine. For this reason, switching the position of specific sidechains in the core can lead to a switch in oligomer states between dimers, trimers, and tetramers in otherwise identical sequences (Harbury et al. 1993). These preferences may, however, not be sufficient to specify the oligomeric state in some cases (Armstrong et al. 2009; Zaytsev et al. 2010). The size of sidechains at core positions also plays a role in establishing the oligomeric state. Whereas the aliphatic residues are compatible with all forms, coiled coils with phenylalanine or tryptophan at all core positions form pentamers, although the presence of a single methionine in the core reduces the phenylalanine pentamer to a tetramer (Liu et al. 2004, 2006a). The oligomeric state can be further influenced by polar residues in the core. Thus, asparagine in position *a* is a determinant of dimeric structure in leucine zippers (Harbury et al. 1993; Gonzalez et al. 1996; Akey et al. 2001), but in position *d* it specifies trimerization in a range of surface proteins (Hartmann et al. 2009). As a general rule, polar and charged residues favour dimers, where they can be more easily solvated, but exceptions abound.

Next to the core residues, the residues in the positions flanking the core (*e* and *g*) also influence oligomerization and orientation. Specifically, if they include a preponderance of hydrophobic residues in one of the positions, they favour tetramers over dimers or trimers, frequently in antiparallel orientation. The resulting, broader hydrophobic core can be described as having two seams of core residues, which overlap by one position, i.e. *g–d* and *d–a*, or *d–a* and *a–e* (Walshaw et al. 2001; Woolfson et al. 2012). Axially symmetric packing of such a larger interface causes the shared position to point towards the central axis (*x* geometry) and the two other positions to point side-ways, enclosing a central cavity (*da* geometry). We have already encountered this packing in Sects. 4.2.3 and 4.2.4, in the context of complementary *x–da* interactions, and of stutters and stammers.

As the two seams move further apart on the face of the helices, they favour increasingly higher oligomers, marking a transition from fibres to tubes (see also Sect. 4.4.2). Thus, adjacent seams (*g-d* and *a-e*) lead to the formation of pentamers, hexamers, and heptamers, and seams separated by an intervening residue (*g-d* and *e-b*, separated by *a;* or *c-g* and *a-e*, separated by *d*) to even higher oligomers, the largest of which is the antiparallel barrel of 12 helices in the multidrug efflux protein TolC (Koronakis et al. 2000). Helices with two seams have been called "bifaceted" and the three types have been denoted I, II, and III in the order of increasing separation (Walshaw et al. 2001; Woolfson et al. 2012). Thomson et al. (2014) have

recently developed a quantitative description of type II bifaceted helices, which has allowed them to engineer penta- to heptameric barrels by computational design. Egelman et al. (2015) have gone one step further by designing a type III bifaceted helix that assembles into sheets (Table 4.1). This open-ended architecture is brought about by two factors: (a) a heterotypic association between the *g–d* seam of one helix and the *e–b* seam of the next, such that the growing sheet always presents a *g–d* seam at one edge and an *e–b* seam at the other; and (b) the placement of ala-nines in position *g* of the *g–d* seam and position *b* of the *e–b* seam, so that a parallel Alacoil interaction can form. Since, in Alacoils, the two helices approach each other more closely along the ridge of alanines than along the other ridge of residues, the angle between the two seams (theoretically 155°) is widened to approximately 180°, allowing the sheet to remain open.

The interface residues thus not only determine the oligomerization and orienta-tion of coiled-coil helices, they also set the geometry of packing interactions (canonical, Alacoil, partially rotated, fully *x–da*). In doing so they navigate an energy landscape in which the different structural states are often nearly isoener-getic and separated by low energy barriers. In our discussion of structural diversity in coiled coils (Sect. 4.4.1), we will review in detail how minor changes to the sequence of the GCN4 leucine zipper can lead to a range of different forms; here we would just like to point out a side-effect of this flat energy landscape, namely that coiled-coil fragments crystallized outside their native protein context often assume non-physiological structures. Instances of this abound in the Protein Structure Databank. For example: (a) the S-helix of the beta-1 subunit of a soluble guanylyl cyclase (3HLS) forms an antiparallel homotetramer, while the parent protein is a parallel heterodimer (with a minor homodimeric form); (b) the pilin subunit of *Neisseria gonorrhoeae* (1AY2) forms an antiparallel dimer, while the parent struc-ture is an extended spiral of offset, parallel subunits; (c) a fragment of the SARS coronavirus spike glycoprotein forms an antiparallel tetramer with *x–da* packing (1ZV7), while the parent protein is a parallel trimer with canonical packing; (d) the cytosolic coiled-coil segments of ion channels have been variously solved as paral-lel tetramers (3BJ4, 2OVC), antiparallel tetramers (3E7K), and parallel trimers (3HFE, 2PNV), even though the parent proteins are all parallel tetramers. A fair number of other such cases could be listed here.

4.3.2   Folding and Stability

Due to the regularity of their interactions, coiled coils are often very stable proteins. Indeed, the most stable protein reported to date may well be the surface-layer pro-tein of the archaeon *Staphylothermus marinus*, whose stalk domain – a homotetra-meric coiled coil – withstands heating to 130 °C in 6 M guanidinium hydrochloride and requires 70 % sulphuric acid for denaturation (Peters et al. 1995). Since this treatment leads to hydrolysis of the peptide bonds, tetrabrachion may be the only known protein where the primary structure appears to be less stable than the

secondary, tertiary, and quaternary structures. This is certainly exceptional, but many coiled coils have been reported to withstand extreme chemical and thermal conditions, including some comprising only a few heptads, such as variants of the GCN4 leucine zipper. The unusual stability that can be achieved by natural coiled coils has been recently reproduced by computational design in engineered coiled coils (Huang et al. 2014). Factors that promote stability are the propensity of the constituent helices to adopt a helical structure, the tightness of their packing interactions, the hydrophobicity of the resulting core, and a shell of favourable polar and ionic interactions that shield this core from solvent. It follows that, all things being equal, the stability of coiled coils should increase with the number of helices, as this provides them with a broader hydrophobic core (at least until the number of helices turns them from bundles into barrels and causes the formation of a central, solvent-filled pore). Indeed, this increase in stability can be observed between the dimeric, trimeric, and tetrameric variants of the GCN4 leucine zipper (Harbury et al. 1993).

Given their thermodynamic stability, it may come as a surprise that coiled coils are close to the unfolded state; in fact, it is not uncommon for them to be mistaken for natively unstructured polypeptides by disorder prediction programs. We are not aware of any systematic study of this, but we have observed it ourselves in many cases over the years, for example with the myosin heavy chain, whose extended stalk is both predicted confidently as a coiled coil and seen as intrinsically disordered by the respective prediction programs (Fig. 4.6). This failure cuts both ways: highly charged sequences that are largely devoid of hydrophobic residues and lack sequence repeats indicative of coiled-coil structure are often predicted as coiled coils, even though they are most likely unstructured (Gruber et al. 2006). This is because coiled coils are highly repetitive, largely solvent-exposed structures and therefore have reduced sequence complexity and a low proportion of hydrophobic to hydrophilic residues relative to globular proteins - just like many natively unstructured sequences. The five most frequent residues in coiled coils (E, L, K, A, Q) comprise more than half of the total (with the charged residues E and K alone constituting more than a quarter), whereas they represent less than a third of residues in globular proteins. Conversely, the five rarest residues in coiled coils (F, H, C, W, P) barely add up to 2 %, whereas they constitute more than 15 % in globular proteins. The similarity between coiled coils and natively unstructured sequences offers a substantial challenge in structure prediction. We find a number of questions helpful in discriminating between the two: (1) Is the sequence repetitive and, if yes, is the periodicity compatible with coiled-coil structure? (2) Does the repetition entail a preponderance of hydrophobic residues at core positions and hydrophilic residues at all other positions? (3) Is the structural context indicative of, or at least compatible with, a coiled coil (for example, does the sequence belong to a protein family that is known to be oligomeric and fibrous)? (4) Is an elevated coiled-coil propensity conserved in homologues? Answering Yes to these questions progressively increases the confidence that a given sequence forms a coiled coil.

In our opinion, coiled-coil sequences have evolved to resemble unstructured polypeptides because they need to ensure in-register folding of rods that are sometimes many hundreds of residues long. Since packing interactions are structurally
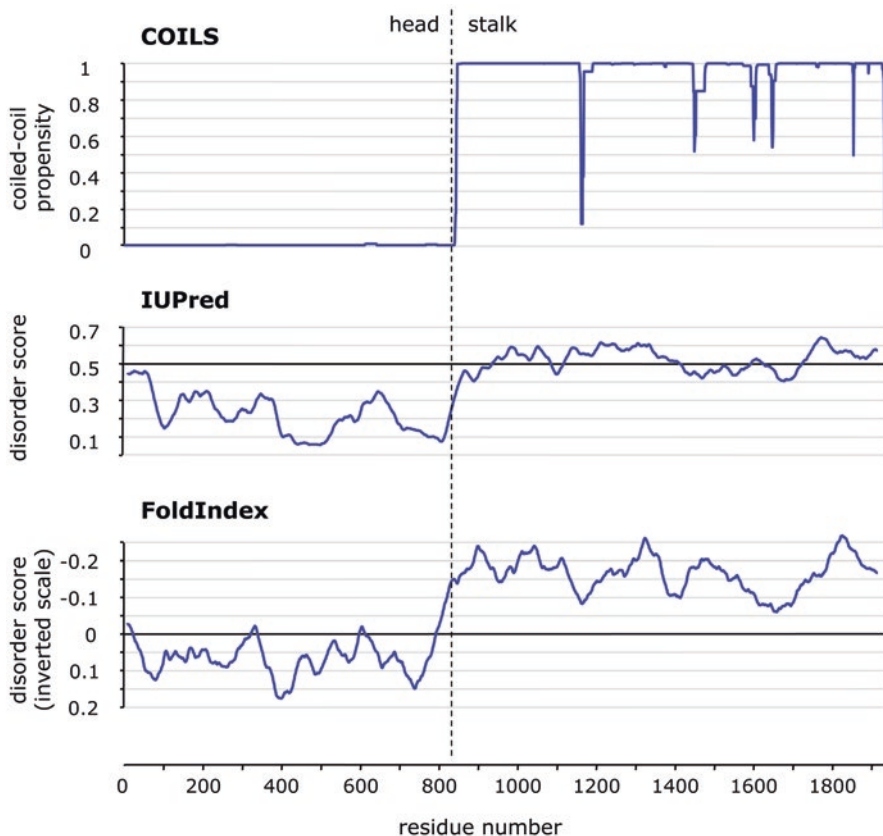
Fig. 4.6 Predictions of coiled-coil propensity and disorder in human myosin heavy chain. The boundary between the globular head domain and the fibrous stalk is marked by a vertical *dotted line*. The output of FoldIndex is shown on an inverted scale in order to make it directly comparable to the two other programs. The graphs show that the rod is recognized both as a coiled coil (COILS) and as natively unstructured (IUPred, Dosztányi et al. 2005; FoldIndex, Prilusky et al. 2005)

the same all along the rod, coiled coils are confronted with many, essentially isoenergetic intermediates that could trap the folding chains out of register if they formed spontaneously. To prevent this, coiled coils have evolved sequences that allow them to be quite stable thermodynamically, once folded, but have kinetic folding barriers that maintain them in an unstructured state until folding has been initiated at a nucleation site and is therefore guaranteed to be in register. Thus, fragments of natural coiled coils, even of considerable size, often do not fold: for example, myosin rod fragments hundreds of residues long remain soluble, but unstructured, if they do not include a nucleation site (Trybus et al. 1997). The concept of nucleation sites as initiators of coiled-coil folding was pioneered by Kammerer, Steinmetz and co-workers, who called them 'trigger sequences' (Steinmetz et al. 1998; Kammerer et al. 1998) and characterized them in a number of different coiled coils. Subsequent

identifications of further trigger sequences (Frank et al. 2000; Wu et al. 2000; Alfadhli et al. 2002; Araya et al. 2002) showed that these nucleation sites do not adhere to a particular consensus sequence, but rather represent short segments of high α-helical propensity, capable of forming many stabilizing interactions in the correct oligomeric form through optimized electrostatic interactions and hydrophobic packing (Lee et al. 2001).

4.4   Structural Diversity

4.4.1   Fibres and Zippers

For decades after their initial description, the concept of coiled coils was closely associated with long fibres, such as the keratins, myosins, epidermins, and fibrinogens of the k-m-e-f class, which were all two- or three-helical. Their properties had greatly helped the original fibre diffraction studies, but caused major impediments to their analysis by X-ray crystallography. Today, there is still only one structure of a k-m-e-f class protein known to high resolution (fibrinogen at 2.7 Å, 1M1J; Yang et al. 2001), while information on the others is available for fragments only or at considerably lower resolution, such as for tropomyosin at 7.0 Å (Whitby and Phillips 2000).

As the X-ray crystallographic analysis of proteins developed, it emerged gradually that coiled-coil interactions were also observed in much shorter helical bundles, often embedded within globular proteins, such as in the catabolite gene activator protein (McKay and Steitz 1981; Nilges and Brunger 1991). These observations led to the proposal that an understanding of coiled-coil structure would be of substantial significance not only for long fibres, but also for globular and membrane proteins, as well as for structures with larger numbers of helices, such as four-helical bundles (Cohen and Parry 1986). Soon afterwards, the discovery of the leucine zipper and particularly of its prototypical representative in the yeast transcription factor GCN4 (Landschultz et al. 1988) provided a tractable model system of great biological importance for the high-resolution analysis and manipulation of coiled coils (O'Shea et al. 1991; Harbury et al. 1993). Today it is widely recognized that coiled coils not only comprise the long fibrous proteins that form filaments, tethers, stalks, levers, and large, mechanically rigid assemblies (e.g. hair, feathers, horn, blood clots), but also many shorter structures, which – among other activities – mediate oligomerization, transduce signals, and facilitate the transport of small molecules (see also Chap. 3 by Hartmann 2017, in this issue).

As an example for the structural diversity accessible to coiled coils we will discuss here the GCN4 leucine zipper, which can be converted to a broad range of structural forms by minor changes to its sequence, illuminating the versatility of the fold and the closeness of its different variants in the energy landscape (Table 4.2). In the native state, it forms a parallel, two-helical coiled coil of 30 residues (GCN4-p1), whose dimerization depends largely on an asparagine residue in

Table 4.2 Diversity of GCN4-derived structures

Name	Oligomer- ization	Orien- tation	Assembly	Core	PDB id	References	Sequence					
							XR	MKQLEDK	VEELLSK	NYHLENE	VARLKKL	VGER
GCN4-p1	Dimer	Parallel	Homo	a–d	2ZTA	O'Shea et al. (1991)	MK	V......	....	....	....	VGER
GCN4-pR	Dimer	Parallel	Homo	a–d			..	....	....	....	....	....
GCN4-pIL	Dimer	Parallel	Homo	a–d	–	Harbury et al. (1993)	..	...I...	I..L...	I..L...	I..L...	I....
GCN4-pII	Trimer	Parallel	Homo	a–d	1GCM	Harbury et al. (1993)	..	...I...	I..I...	I..I...	I..I...	I....
GCN4-pLI	Tetramer	Parallel	Homo	a–d	1GCL	Harbury et al. (1993)	..	...I...	L..I...	L..I...	L..I...	L....
GCN4-pVLS	Trimer	Parallel	Homo	a–d	1IJ0	Akey et al. (2001)	..	....	...S...	V......	....	....
GCN4-pVLT	Trimer	Parallel	Homo	a–d	1IJ1	Akey et al. (2001)	..	....	...T...	V......	....	....
GABH	Dimer	Parallel	Hetero	a–d	1KD8	Keating et al. (2001)	.E	V....AE	...IE.E	VW.....	...E.E	NA.CEA
							.K	V...KA.	....K..	.LW..K.K	.......K	NA.CKA
GCN4-p1 N16A	Trimer	Antiparallel	Homo	a–d	1RB1	Holton and Alber (2004)	-.	....	....	A......	....	....
GCN4-pM3	Trimer	Parallel	Homo	a–d	2O7H	Ciani et al. (2010)	S.	....	....	......R	...E..	....
GCN4-pLI E20C	Tetramer	Antiparallel	Homo	x–da	2CCN	Yadav et al. (2006)	..	...I...	L..I...	L..IC..	L..I...	L....
GCN4-pLI E20S	Tetramer	Parallel	Homo	a–d	2CCE	Yadav et al. (2006)	..	...I...	L..I...	L..IS.	L..I...	L....
GCN4-pLI E20S	Tetramer	Antiparallel	Homo	x–da	2CCF	Yadav et al. (2006)	..	...I...	L..I...	L..IS.	L..I...	L....

4   Coiled-coil Structure   119

GCN4-pAg	Tetramer	Antiparallel	Homo	Alacoil	2B1F	Deng et al. (2006)	MK	V.....A	.......A	.......A	....	....
GCN4-pAe$_{Lv}$	Tetramer	Antiparallel	Homo	x–da	2R2V	Liu et al. (2007)	MK	L..VA..	L..VA..	L..NA..	L..VA..	L....
GCN4-pVg	Tetramer	Antiparallel	Homo	x–da	2B22	Deng et al. (2006)	MK	V......V	V	V	....	....
GCN4-pVe	Tetramer	Parallel	Homo	offset^a	2IPZ	Liu et al. (2006a)	MK	V...V..	....	...V..	...V..	....
GCN4-pAA	Heptamer	Parallel	Homo	x–da	2HY6	Liu et al. (2006b)	MK	V...A.A	...A.A	...A.A	...A.A	....A.A
GCN4-pVe/	Tetramer	Antiparallel	Hetero	x–da	3CK4	Deng et al. (2008)	MK	V...V..	....	...V..	...V..	....
CN4-pVg							MK	V......V	V	V	....	....
GCN4-pAe/	Tetramer	Antiparallel	Hetero	x–da	3CRP	Deng et al. (2008)	MK	V...A..	...A..	...A..	...A..	....
GCN4-pAg							MK	V......A	.......A	.......A	....	....

^{a}Helices A and B are in the same register, C is shifted axially by about 3.6 A and D by about 5.8 A, leading to the formation of mixed a–a–d–d and d–d–a–a core layers. The diagonally opposed helices A&C are much closer together than B&D, giving the bundle a trapezoid cross section. The bundle has an unusual core geometry, with helix A drifting from an a–d–e core at its N-terminus to an intermediate rotation state at its C-terminus, helix B remaining fully canonical, helix C assuming an a–d–e core throughout, and helix D assuming the same intermediate rotation state as helix A does at its C-terminus

position *a* of the heptad repeat (N16). This residue represents an instance of 'negative design' since it provides structural specificity at the expense of stability; its replacement generally increases stability, but leads to trimers, or mixtures of dimers and trimers (Harbury et al. 1993; Gonzalez et al. 1996; Akey et al. 2001). All these forms are parallel, but in one case an antiparallel trimer was obtained by mutating N16 to alanine. This mutation caused the structure to become trimeric and invert the orientation of one helix, such that the A16 residues form mixed *a-d* core layers with L12, filling the cavity caused by the small alanine sidechains. The parallel orientation of the helices could be rescued by adding benzene to the solvent, which filled the hydrophobic cavity in the core (Holton and Albert 2004). The oligomeric state of GCN4 could also be altered by other substitutions in the hydrophobic core: as outlined in Sect. 4.3.1, the geometry of packing interactions leads coiled-coil dimers to favour β-branched residues in position *a* and γ-branched or unbranched residues in position *d*, and tetramers to favour the reverse; trimers show no preference. This rule was, in fact, derived from GCN4 variants, for which isoleucines in *a* and leucines in *d* produced dimers (GCN4-IL) and the reverse distribution tetramers (GCN4-LI; Harbury et al. 1993). A retro-GCN4, consisting of an inversion of the GCN4 sequence, therefore unsurprisingly yielded a tetramer (Mittl et al. 2000). Other combinations of isoleucine, leucine, and valine typically resulted in trimers or, rarely, mixtures of dimers and trimers (Harbury et al. 1994).

Structural diversity was also brought about by mutating positions *e* and *g*, flanking the hydrophobic core (Table 4.2). For example, introducing an ionic bond between neighbouring chains via an arginine in position *g* and a glutamate in position *e* overrode the the effect of N16 and converted GCN4-p1 to a trimer, but only when this was done at the trigger site of the coiled coil (Ciani et al. 2010). A more comprehensive set of polar mutations in positions *e* and *g*, aimed at obtaining negatively and positively charged variants of the helices, resulted in heterodimers between these (Keating et al. 2001). Most other mutations in the flanking positions, however, involved an increase in hydrophobicity and generally produced tetramers, albeit with great heterogeneity in orientation and core packing geometry (Table 4.2). Thus, mutating all polar residues in position *e* to valine (GCN4-pVe) yielded an odd, parallel four-helical bundle with offset helices that engage in a mixture of canonical and complementary *x–da* interactions, whereas the equivalent mutation of polar residues in position *g* (GCN4-pVg) yielded a symmetrical, antiparallel tetramer with near-ideal *x–da* packing (Liu et al. 2006b; Deng et al. 2006; Dunin-Horkawicz and Lupas 2010a). A similar replacement of polar residues in position *g* with alanine (GCN4-pAg) also yielded an antiparallel tetramer, but the axial rotation of the helices went so far that a new canonical core was formed between positions *d* and *g*, resulting in the only GCN4 variant with Alacoil packing (Deng et al. 2006; Table 4.1). As we discussed in Sect. 4.2.3, placing small, hydrophobic residues in positions *e* or *g* biases the helices towards complementary *x–da* packing with an *a–d–e* or *a–d–g* core, respectively. Deng et al. (2008) exploited this in order to obtain heteromeric antiparallel tetramers with mixed *a–d–g* and *a–d–e* cores by combining either GCN4-pVe with GCN4-pVg, or the equivalent variants carrying alanine in place of valine. Combining hydrophobic mutations in *e* and *g* into the

same helix (GCN4-pAA) yielded the most unusual zipper variant yet: a heptameric coiled-coil tube with the helices staggered in phase with the heptad repeat (Liu et al. 2006c).

The diversity of GCN4 variants highlights how close these are in their overall energy. The fact that the N16A mutant can be switched from an antiparallel to a parallel trimer by the addition of benzene is evidence of this, as is the observation of Yadav et al. (2006) that the tetrameric GCN4 variant, GCN4-pLI, when carrying a single point mutation from glutamate to serine in position *e*, can be switched between a parallel form with canonical packing and an antiparallel form with *x–da* packing (Table 4.2), simply by varying the buffer. Collectively, GCN4 variants map out how simple changes can lead from a plain dimeric fibre to the higher-order assemblies seen in many proteins.

4.4.2 Tubes, Sheets, Spirals, and Rings

As discussed in Sect. 4.3.1, coiled coils can progressively increase the number of helices by broadening their hydrophobic core, then splitting it into two adjacent cores. Starting with four-helical bundles, the space along their central axis increases to form cavities, frequently containing water or other solvent molecules; from five-helical bundles onward, the cavities merge into a continuous, solvent-filled pore. The resulting structures have been referred to variously as tubes (Lupas and Gruber 2005), α-barrels (Koronakis et al. 2000), or α-cylinders (Walshaw et al. 2001; Walshaw and Woolfson 2003). As the diameter of the pore widens and packing interactions become more irregular, the constituent helices may also start to deviate from the perpendicular, yielding funnel-shaped tubes, for example in the upper collar protein of the Bacillus bacteriophage φ29 (Simpson et al. 2000). The potential of coiled-coil tubes to mediate solute transport when embedded in the membrane has been recognized for at least two decades (Malashkevich et al. 1996) and recently Joh et al. (2014) designed the first synthetic ion antiporter by exploiting the axial cavities of a membrane-embedded four-helical coiled coil. The differential chemical accessibility of the central channel in aqueous environment has also suggested a potential to engineer new enzymes (Burton et al. 2013). So far, tubes with five, six, ten, and twelve helices have been identified in nature, and this range has been complemented by computational design to include engineered five-, six-, and seven-helical structures (Thomson et al. 2014; Huang et al. 2014).

When the interaction seams of bifaceted helices are maximally separated, they may come to specify angles close to 180°, depending on the size distribution of residues at the interfaces. In such cases, the helices assemble into open α-sheets, rather than into circular tubes. Similarly, pairs of helices that neutralize each other's angular offset, such as in the archaeal protein pT26-6p (2WB7; Table 4.1), also lead to open structures (Walshaw et al. 2001; Walshaw and Woolfson 2003). Most natural examples consist of three or four helices, but recently, computational design has resulted in sheets with a much higher number of helices, as discussed in Sect. 4.3.1;

four of these sheets assemble to form a novel type of fibre (Egelman et al. 2015) (Fig. 4.7). Maintaining knobs-into-holes contacts in a sheet requires all but the edge helices to supercoil in opposite directions. The strain resulting from this conformational conflict distorts the packing interactions and gradually moves the helices out of register, restricting the length of these associations to typically less than four heptads.

Sheets meet tubes in the formation of spirals. Compared to tubes, the helices of spirals are noticeably offset from each other along the axis of the coiled coil. For example, the major coat proteins of filamentous bacteriophages form staggered and slightly curved sheets, which then assemble into a multi-stranded spiral (Fig. 4.7), an architecture also found in proteins of the bacterial flagella and pili (reviewed in more detail in Lupas and Gruber 2005). Another example is the multidrug resistance protein MexA, which was crystallized as a complex, tail-to-tail assembly of two spirals, one with six and the other with seven subunits. Each subunit contributes a helical hairpin to the spiral and the interaction between adjacent hairpins is set off by one heptad. This structure probably represents another instance of the crystallization artefacts discussed in Sect. 4.3.1, as MexA most likely assembles into a funnel *in vivo* (Symmons et al. 2009). Some of the GCN4 variations that we have discussed in Sect. 4.4.1 are staggered as well, most notably the heptameric GCN4-pAA. The mechanism, however, differs. Spirals make knobs-into-holes contacts between core residues in different heptads, whereas these coiled coils make homotypic contacts but are axially shifted by the equivalent of one residue per helix (Liu et al. 2006c).

An even more intriguing case is that of the apolipoproteins, which, despite being clearly homologous, form structures differing substantially in packing, supercoiling, and oligomer state. Apolipoproteins A-I and A-II oligomerize into ring-shaped four-helical bundles with an underlying 22/6 periodicity (Fig. 4.7), but show only local knobs-into-holes packing and supercoil angles similar to undistorted helices. In contrast, apolipoprotein E and apolipophorin III form monomeric left-handed coiled coils with regular knobs-into-holes interactions (Boguski et al. 1985; reviewed in detail in Lupas and Gruber 2005).

4.5   Evolution and Phylogenetic Diversity

The large diversity of coiled-coil structures discussed above begs the question of their evolutionary origins. Some clearly form ancient families, which can occasionally be traced as far as the root of a kingdom (witness the leucine zipper transcription factors of eukaryotes), but in most cases sequence similarity searches do not uncover homologues beyond individual phyla and usually not even that far. Only in exceptional cases can a coiled coil be traced back all the way to the time of the Last Universal Common Ancestor (e.g. in seryl-tRNA synthetase). We would like to offer two reasons why this might be the case. One is that coiled coils diverge faster than most proteins, due to their primarily structural function, which has lower
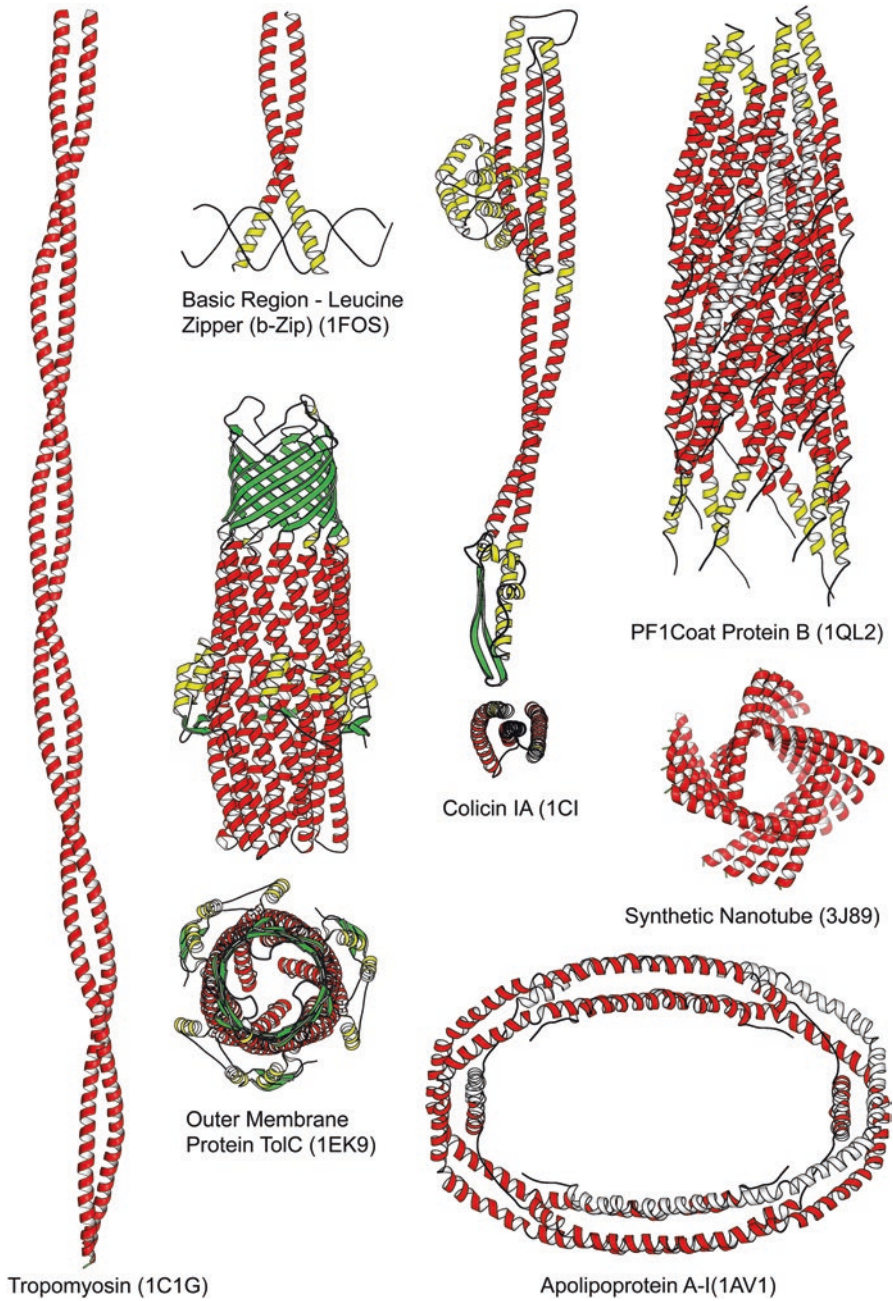
Fig. 4.7 Diversity of coiled-coil structures. The figure shows a fibre (tropomyosin), a zipper (the Fos b-Zip domain bound to DNA), a tube in *side* and *top view* (TolC), a sheet in *side* and *top view* (colicin IA), a spiral (phage PF1 coat protein B), a synthetic nanotube assembled from sheets, and a ring (apolipoprotein A-I)

constraints than catalytic activity, and their highly repetitive nature, which makes them more resilient against point mutations than other structures. At the same time, they converge faster than most other proteins, due to their repetitive structure, which favours the same patterns of hydrophobic and hydrophilic residues, and the same restricted alphabet of residues with high helical propensity. For these reasons, the point in the past where it becomes impossible to distinguish homology from analogy in statistical sequence analyses is more recent for coiled coils than for other proteins. This does not mean that individual coiled-coil families are not older than this point, it just means that we can't trace their origins beyond it. Additionally, as discussed in Sect. 4.3.2, coiled coils seem to have been under evolutionary pressure to evolve towards unstructured sequences for reasons of folding specificity, which has further lowered their sequence complexity and increased their rate of divergence. In this context, we note the existence of coiled coils with highly biased residue distributions, for example in a family of proteins we have analyzed in great detail, the trimeric autotransporter adhesins (TAAs; Szczesny and Lupas 2008; Bassler et al. 2015); here, one can find proteins such as Bcep18194_B0441 of *Burkholderia lata*, which has two extensive coiled-coil segments of ~700 and ~1200 residues, respectively, consisting to two-thirds of serine and threonine, or the putative adhesin VEIDISOL_00919 of *Veillonella dispar*, in whose stalk of more than 2000 residues a quarter are asparagine. Other examples of this kind can be readily found, also in eukaryotic proteins.

The second reason why coiled-coil domains might be difficult to trace far back in time is their rapid turnover in many families. Like many other repetitive structures, they seem to constantly evolve by amplification and divergence from individual repeats but, unlike most other repetitive structures, their repeat unit is so short that it is easy to recruit from sequences not part of a coiled coil, or even from non-coding sequence. Coiled coils thus seem to readily evolve *de novo*. As evidence for this we see coiled coils, primarily in prokaryotes, whose repeat units are essentially identical to each other over hundreds of residues, not only in the protein sequence, but also at the level of the DNA; such coiled coils typically occur only in a few closely related species and are therefore probably not more than a few million years old. We see further evidence for *de novo* evolution in the many protein families that are generally globular, but in which one or a few members have "grown" coiled coils through the extension of helices that are an integral part of the family fold (Lupas and Gruber 2005). Although more difficult to establish, it seems reasonable to assume that the *de novo* acquisition of coiled coils is paralleled by a corresponding rate of loss.

These considerations, of course, proceed from sequences that we can recognize as coiled coils, either because they are considered to assume this structure by prediction programs, or (much more rarely) because we have structural information. Not considered are an unknown number of sequences that form coiled coils, but remain unpredicted because of an odd residue composition or an odd periodicity. The surprising recent discovery of α/β coiled coils (Hartmann et al. 2016) as a substantially new variant of a fold we thought we understood comprehensively suggests that there might be yet more to this seemingly simple fold than we realize.

References

Akey DL, Malashkevich VN, Kim PS (2001) Buried polar residues in coiled-coil interfaces. Biochemistry 40:6352–6360

Alfadhli A, Steel E, Finlay L, Bächinger HP, Barklis E (2002) Hantavirus nucleocapsid protein coiled-coil domains. J Biol Chem 277:27103–27108

Araya E, Berthier C, Kim E, Yeung T, Wang X, Helfman DM (2002) Regulation of coiled-coil assembly in tropomyosins. J Struct Biol 137:176–183

Armstrong CT, Boyle AL, Bromley EH, Mahmoud ZN, Smith L, Thomson AR, Woolfson DN (2009) Rational design of peptide-based building blocks for nanoscience and synthetic biology. Faraday Discuss 143:305–317

Armstrong CT, Vincent TL, Green PJ, Woolfson DN (2011) SCORER 2.0: an algorithm for distinguishing parallel dimeric and trimeric coiled-coil sequences. Bioinformatics 27:1908–1914

Astbury, W. T. (1938). The fourth Spiers memorial lecture. X-ray adventures among the proteins. Transactions of the Faraday Society 34, 378–388.

Astbury WT (1946) X-ray studies of nucleic acids. Sym Soc Exp Biol (1):66–76

Bassler J, Alvarez BH, Hartmann MD, Lupas AN (2015) A domain dictionary of trimeric autotransporter adhesins. Int J Med Microbiol 305:265–275

Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS (1995) Predicting coiled coils by use of pairwise residue correlations. Proc Natl Acad Sci 92:8259–8263

Boguski MS, Elshourbagy N, Taylor JM, Gordon JI (1985) Comparative analysis of repeated sequences in rat apolipoproteins AI, A-IV, and E. Proc Natl Acad Sci 82:992–996

Braun V, Bosch V (1972) Repetitive sequences in the murein-lipoprotein of the cell wall of *Escherichia coli*. Proc Natl Acad Sci 69:970–974

Burkhard P, Stetefeld J, Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. Trends Cell Biol 11:82–88

Burton AJ, Thomas F, Agnew C, Hudson KL, Halford SE, Brady RL, Woolfson DN (2013) Accessibility, reactivity, and selectivity of side chains within a channel of de novo peptide assembly. J Am Chem Soc 135:12524–12527

Chothia C, Levitt M, Richardson D (1977) Structure of proteins: packing of α-helices and pleated sheets. Proc Natl Acad Sci 74:4130–4134

Chothia C, Levitt M, Richardson D (1981) Helix to helix packing in proteins. J Mol Biol 145:215–250

Ciani B, Bjelic S, Honnappa S, Jawhari H, Jaussi R, Payapilly A, Jowitt T, Steinmetz MO, Kammerer RA (2010) Molecular basis of coiled-coil oligomerization-state specificity. Proc Natl Acad Sci 107:19850–19855

Cohen C (1998) Why fibrous proteins are romantic. J Struct Biol 122:3–16

Cohen C, Parry DAD (1986) α-helical coiled coils - a widespread motif in proteins. Trends Biochem Sci 11:245–248

Cohen C, Parry DAD (1990) α-helical coiled coils and bundles: how to design an α-helical protein. Proteins 7:1–15

Cohen C, Parry DAD (1994) Alpha-Helical Coiled Coils-More Facts and Better Predictions. Science 263:488–489

Crick FHC (1952) Is α-keratin a coiled coil? Nature 170:882–883

Crick FHC (1953a) The Fourier transform of a coiled-coil. Acta Crystallogr 6:685–689

Crick FHC (1953b) The packing of α-helices: simple coiled-coils. Acta Crystallogr 6:689–697

Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617–625

Deng Y, Liu J, Zheng Q, Eliezer D, Kallenbach NR, Lu M (2006) Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. Structure 14:247–255

Deng Y, Liu J, Zheng Q, Li Q, Kallenbach NR, Lu M (2008) A Heterospecific Leucine Zipper Tetramer. Chem Biol 15:908–919

Dosztányi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21:3433–3434

Dunin-Horkawicz S, Lupas AN (2010a) Measuring the conformational space of square four-helical bundles with the program samCC. J Struct Biol 170:226–235

Dunin-Horkawicz S, Lupas AN (2010b) Comprehensive analysis of HAMP domains: implications for transmembrane signal transduction. J Mol Biol 397:1156–1174

Egelman EH, Xu C, DiMaio F, Magnotti E, Modlin C, Yu X, Wright E, Baker D, Conticello VP (2015) Structural plasticity of helical nanotubes based on coiled-coil assemblies. Structure 23:280–289

Ferris HU, Dunin-Horkawicz S, Mondéjar LG, Hulko M, Hantke K, Martin J, Schultz JE, Zeth K, Lupas AN, Coles M (2011) The mechanisms of HAMP-mediated signaling in transmembrane receptors. Structure 19:378–385

Ferris HU, Dunin-Horkawicz S, Hornig N, Hulko M, Martin J, Schultz JE, Zeth K, Lupas AN, Coles M (2012) Mechanism of regulation of receptor histidine kinases. Structure 20:56–66

Ferris HU, Zeth K, Hulko M, Dunin-Horkawicz S, Lupas AN (2014) Axial helix rotation as a mechanism for signal regulation inferred from the crystallographic analysis of the *E. coli* serine chemoreceptor. J Struct Biol 186:349–356

Frank S, Lustig A, Schulthess T, Engel J, Kammerer RA (2000) A distinct seven-residue trigger sequence is indispensable for proper coiled-coil formation of the human macrophage scavenger receptor oligomerization domain. J Biol Chem 275:11672–11677

Gernert KM, Surles MC, Labean TH, Richardson JS, Richardson DC (1995) The Alacoil: a very tight, antiparallel coiled-coil of helices. Protein Sci 4:2252–2260

Gonzalez L Jr, Woolfson DN, Alber T (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat Struct Mol Biol 3:1011–1018

Grigoryan G, DeGrado WF (2011) Probing designability via a generalized model of helical bundle geometry. J Mol Biol 405:1079–1100

Gruber M, Lupas AN (2003) Historical review: another 50th anniversary - new periodicities in coiled coils. Trends Biochem Sci 28:679–685

Gruber M, Söding J, Lupas AN (2006) Comparative analysis of coiled-coil prediction methods. J Struct Biol 155:140–145

Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262:1401–1407

Harbury PB, Kim P, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. Nature 371:80–83

Hartmann MD (2017) Functional and structural roles of coiled coils. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Hartmann MD, Ridderbusch O, Zeth K, Albrecht R, Testa O, Woolfson DN, Sauer G, Dunin-Horkawicz S, Lupas AN, Alvarez BH (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. Proc Natl Acad Sci 106:16950–16955

Hartmann MD, Mendler CT, Bassler J, Karamichali I, Ridderbusch O, Lupas AN, Hernandez Alvarez B (2016) α/β coiled coils. *eLife* 5:e11861

Hicks MR, Walshaw J, Woolfson DN (2002) Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. J Struct Biol 137:73–81

Holton J, Alber T (2004) Automated protein crystal structure determination using ELVES. Proc Natl Acad Sci 101:1537–1542

Huang PS, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B, Baker D (2014) High thermodynamic stability of parametrically designed helical bundles. Science 346:481–485

Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, Schultz A, Martin J, Schultz JE, Lupas AN, Coles M (2006) The HAMP domain structure implies helix rotation in transmembrane signaling. Cell 126:929–940

Joh NH, Wang T, Bhate MP, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G, DeGrado WF (2014) De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. Science 346:1520–1524

Judson HF (1979) The eighth day of creation. Simon & Schuster, New York

Kammerer RA, Schulthess T, Landwehr R, Lustig A, Engel J, Aebi U, Steinmetz MO (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils. Proc Natl Acad Sci 95:13419–13424

Keating AE, Malashkevich VN, Tidor B, Kim PS (2001) Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. Proc Natl Acad Sci 98:14825–14830

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181:662–666

Koretke KK, Szczesny P, Gruber M, Lupas AN (2006) Model structure of the prototypical non-fimbrial adhesin YadA of Yersinia enterocolitica. J Struct Biol 155:154–161

Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature 405:914–919

Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764

Lee DL, Lavigne P, Hodges RS (2001) Are trigger sequences essential in the folding of two-stranded α-helical coiled-coils? J Mol Biol 306:539–553

Liu J, Yong W, Deng Y, Kallenbach NR, Lu M (2004) Atomic structure of a tryptophan-zipper pentamer. Proc Natl Acad Sci 101:16156–16161

Liu J, Zheng Q, Deng Y, Kallenbach NR, Lu M (2006a) Conformational transition between four and five-stranded phenylalanine zippers determined by a local packing interaction. J Mol Biol 361:168–179

Liu J, Deng Y, Zheng Q, Cheng CS, Kallenbach NR, Lu M (2006b) A parallel coiled-coil tetramer with offset helices. Biochemistry 45:15224–15231

Liu J, Zheng Q, Deng Y, Cheng CS, Kallenbach NR, Lu M (2006c) A seven-helix coiled coil. Proc Natl Acad Sci 103:15457–15462

Liu J, Zheng Q, Deng Y, Qunnu L, Kallenbach NR, Lu M (2007) Conformational specificity of the lac repressor coiled-coil tetramerization domain. Biochemistry 46:14951–14959

Lupas A (1996) Coiled coils: new structures and new functions. Trends Biochem Sci 21:375–382

Lupas AN, Gruber M (2005) The structure of α-helical coiled coils. Adv Protein Chem 70:37–78

Lupas A, Van Dyke S, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

Mahrenholz CC, Abfalter IG, Bodenhofer U, Volkmer R, Hochreiter S (2011) Complex networks govern coiled-coil oligomerization–predicting and profiling by means of a machine learning approach. Mol Cell Proteomics 10:M110–004994

Malashkevich VN, Kammerer RA, Efimov VP, Schulthess T, Engel J (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel? Science 274:761–765

Mason JM, Arndt KM (2004) Coiled coil domains: stability, specificity, and biological implications. Chembiochem 5:170–176

McDonnell AV, Jiang T, Keating AE, Berger B (2006) Paircoil2: improved prediction of coiled coils from sequence. Bioinformatics 22:356–358

McKay DB, Steitz T (1981) Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature 290:744–749

McLachlan AD, Stewart M (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J Mol Biol 98:293–304

McLachlan AD, Stewart M (1976) The 14-fold periodicity in α-tropomyosin and the interaction with actin. J Mol Biol 103:271–298

Mittl PRE, Deillon C, Sargent D, Liu N, Klauser S, Thomas RM, Gutte B, Grütter MG (2000) The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure. Proc Natl Acad Sci 97:2562–2566

Mondéjar LG, Lupas A, Schultz A, Schultz JE (2012) HAMP domain-mediated signal transduction probed with a mycobacterial adenylyl cyclases as a reporter. J Biol Chem 287:1022–1031

Nilges M, Brünger AT (1991) Automated modeling of coiled coils: application to the GCN4 dimerization region. Protein Eng 4:649–659

Offer G, Hicks MR, Woolfson DN (2002) Generalized Crick equations for modeling noncanonical coiled coils. J Struct Biol 137:41–53

O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254:539–544

Parry DAD (1975) Analysis of the primary sequence of α-tropomyosin from rabbit skeletal muscle. J Mol Biol 98:519–535

Parry DAD (1982) Coiled-coils in α-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. Biosci Rep 2:1017–1024

Parry DAD, Crewther WG, Fraser RD, MacRae TP (1977) Structure of α-keratin: structural implication of the amino acid sequences of the type I and type II chain segments. J Mol Biol 113:449–454

Parry DAD, Fraser RB, Squire JM (2008) Fifty years of coiled-coils and α-helical bundles: A close relationship between sequence and structure. J Struct Biol 163:258–269

Pauling L, Corey RB (1950) Two hydrogen-bonded spiral configurations of the polypeptide chain. J Am Chem Soc 72:534

Pauling L, Corey RB (1953) Compound helical configurations of polypeptide chains: structure of proteins of the α-keratin type. Nature 171:59–61

Pauling L, Corey RB, Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci 37:205–211

Peters J, Nitsch M, Kuhlmorgen B, Golbik R, Lupas A, Kellermann J, Engelhardt H, Pfander JP, Müller S, Goldie K, Engel A, Stetter K-O, Baumeister W (1995) Tetrabrachion: a filamentous archaebacterial surface protein assembly of unusual structure and extreme stability. J Mol Biol 245:385–401

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435–3438

Simpson AA, Tao Y, Leiman PG, Badasso MO, He Y, Jardine PJ, Olson NH, Morais MC, Grimes S, Anderson DL, Baker TS (2000) Structure of the bacteriophage φ29 DNA packaging motor. Nature 408(6813):745–750

Solan A, Ratia K, Fairman R (2002) Exploring the role of alanine in the structure of the Lac repressor tetramerization domain, a ferritin-like Alacoil. J Mol Biol 317:601–612

Steinmetz MO, Stock A, Schulthess T, Landwehr R, Lustig A, Faix J, Gerisch G, Aebi U, Kammerer RA (1998) A distinct 14 residue site triggers coiled-coil formation in cortexillin I. EMBO J 17:1883–1891

Stone D, Sodek J, Johnson P, Smillie LB (1975) Tropomyosin: correlation of amino acid sequence and structure. FEBS Proc 31:125–136

Strelkov SV, Burkhard P (2002) Analysis of α-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J Struct Biol 137:54–64

Symmons MF, Bokma E, Koronakis E, Hughes C, Koronakis V (2009) The assembled structure of a complete tripartite bacterial multidrug efflux pump. Proc Natl Acad Sci 106:7173–7178

Szczepaniak K, Lach G, Bujnicki JM, Dunin-Horkawicz S (2014) Designability landscape reveals sequence features that define axial helix rotation in four-helical homo-oligomeric antiparallel coiled-coil structures. J Struct Biol 188:123–133

Szczesny P, Lupas A (2008) Domain annotation of trimeric autotransporter adhesins—daTAA. Bioinformatics 24:1251–1256

Testa OD, Moutevelis E, Woolfson DN (2009) CC+: a relational database of coiled-coil structures. Nucleic Acids Res 37:D315–D322

Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble α-helical barrels. Science 346:485–488

Trigg J, Gutwin K, Keating AE, Berger B (2011) Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. PLoS One 6:e23519

Trybus KM, Freyzon Y, Faust LZ, Sweeney HL (1997) Spare the rod, spoil the regulation: necessity for a myosin rod. Proc Natl Acad Sci 94:48–52

Vincent TL, Green PJ, Woolfson DN (2013) LOGICOIL—multi-state prediction of coiled-coil oligomeric state. Bioinformatics 29:69–76

Walshaw J, Woolfson DN (2001) SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures. J Mol Biol 307:1427–1450

Walshaw J, Woolfson DN (2003) Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. J Struct Biol 144:349–361

Walshaw J, Shipway JM, Woolfson DN (2001) Guidelines for the assembly of novel coiled-coil structures: α-sheets and α-cylinders. Biochem Soc Symp 68:111–123

Whitby FG, Phillips GN (2000) Crystal structure of tropomyosin at 7 Ångstroms resolution. Proteins 38:49–59

Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289:366–373

Wolf E, Kim PS, Berger B (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci 6:1179–1189

Wood CW, Bruning M, Ibarra AÁ, Bartlett GJ, Thomson AR, Sessions RB, Brady RL, Woolfson DN (2014) CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinformatics 30:3029–3035

Woolfson DN (2005) The design of coiled-coil structures and assemblies. Adv Protein Chem 70:79–112

Woolfson DN (2017) Coiled-coil design: updated and upgraded. Fibrous proteins: structures and mechanisms. Springer, Dordrecht

Woolfson DN, Alber T (1995) Predicting oligomerization states of coiled coils. Protein Sci 4:1596–1607

Woolfson DN, Bartlett GJ, Bruning M, Thomson AR (2012) New currency for old rope: from coiled-coil assemblies to α-helical barrels. Curr Opin Struct Biol 22:432–441

Wu KC, Bryan JT, Morasso MI, Jang S-I, Lee J-H, Yang J-M, Marekov LN, Parry DAD, Steinert PM (2000) Coiled-coil trigger motifs in the 1B and 2B rod domain segments are required for the stability of keratin intermediate filaments. Mol Biol Cell 11:3539–3558

Yadav MK, Leman LJ, Price DJ, Brooks CL 3rd, Stout CD, Ghadiri MR (2006) Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotet-rameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution. Biochemistry 45:4463–4473

Yang Z, Kollman JM, Pandi L, Doolittle RF (2001) Crystal structure of native chicken fibrinogen at 2.7 Å resolution. Biochemistry 40:12515–12523

Zaytsev DV, Xie F, Mukherjee M, Bludin A, Demeler B, Breece RM, Tierney DL, Ogawa MY (2010) Nanometer to millimeter scale peptide-porphyrin materials. Biomacromolecules 11:2602–2609

Chapter 5
Structural Transition of Trichocyte Keratin Intermediate Filaments During Development in the Hair Follicle

R.D. Bruce Fraser and David A.D. Parry

Contents

Abstract The intermediate filaments (IF) in trichocyte (hard α-) keratin are unique amongst the various classes of IF in having not one but two topologically-distinct structures. The first is formed at an early stage of hair development in a reducing environment within the cells in the lower part of the follicle. The second structure occurs at a later stage of hair development in the upper part of the follicle, where there is a transition to an oxidizing environment. Crosslinking studies reveal that molecular slippage occurs within the IF upon oxidation and that this results in many cysteine residues lying in near axial alignment, thereby facilitating disulphide bond formation. The disulphide bonds so formed stabilize the assembly of IF molecules and convert the keratin fibre into a tough, resilient and insoluble structure suitable for its function *in vivo* as a thermo-regulator and a protector of the animal against its external environment.

R.D.B. Fraser
Institute of Fundamental Sciences, Massey University,
Private Bag 11-222, Palmerston North 4442, New Zealand

28 Satinay Drive, Noosa Parklands, Tewantin, Qld 4565, Australia

D.A.D. Parry (✉)
Institute of Fundamental Sciences and Riddet Institute, Massey University,
Private Bag 11-222, Palmerston North 4442, New Zealand
e-mail: d.parry@massey.ac.nz

Abbreviations

DST   disulpho-succinimidyl tartrate
IF   intermediate filament
HPLC   high-performance liquid chromatography
TEM   transmission electron microscopy
STEM   scanning transmission electron microscopy

5.1   Introduction

Trichocyte keratins, sometimes known as hard α-keratins, are the major protein con-
stituents of epidermal appendages that include hair, wool, hoof, horn, claw, baleen
and quill. Early transmission electron microscope (TEM) studies of cross-sections
showed that the cortex of these keratins is composed of filaments about 7–10 nm in
diameter embedded in an osmiophilic matrix (Fig. 5.1: Birbeck and Mercer 1957;
Rogers 1959; Jones 1976). The filaments were initially referred to as microfibrils
but are now more commonly referred to as "intermediate filaments" (IF) in line with
structures of similar size and sequence characteristics seen widely in eukaryotic
cells. In some cases the filaments observed in the electron microscope were seen to
exhibit a 'ring-core' structure and, in other cases, the filaments displayed only a
'ring' structure, as illustrated in the inset to Fig. 5.1a. Which of these best repre-
sented the *in vivo* situation was a subject of wide debate in the 1960s. As will be
seen later, however, there is a structural rearrangement in the IF accompanying the
transition from the reduced to the oxidized state during the progression of the fibre
up the follicle, and it is not unlikely that the differing electron microscope observa-
tions reflect that change.

   X-ray diffraction patterns have been obtained from a variety of trichocyte kera-
tins, the most detailed so far being from porcupine quill. The patterns have revealed
meridional reflections with spacings of about 0.515 and 0.1485 nm, a strong near-
equatorial layer line with an axial spacing of about 7 nm and an equatorial maxi-
mum with a spacing of about 1 nm (Astbury and Woods 1930, 1933; MacArthur
1943). All of these features were subsequently shown to be features characteristic of
a coiled-coil structure with α-helical strands (Fig. 5.2: Fraser et al. 1964, 1965,
1976). These observations were naturally associated with the oriented portion of the
fibre structure, i.e. the filaments, rather than the matrix.

   Early biochemical studies also showed that the proteins in trichocyte keratins
could be separated into low- and high-sulphur fractions (see, for example, Crewther
et al. 1965; Fraser et al. 1972). The low-sulphur fraction was shown to have a
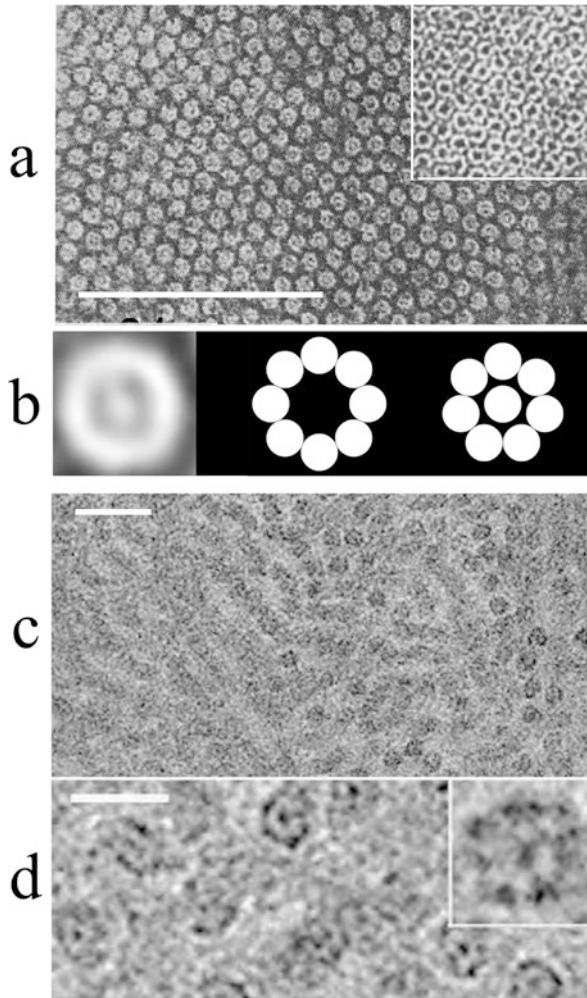
Fig. 5.1 (**a**) Transmission electron micrograph of a cross-section of the macrofibrils in the para-cortex of fine Merino wool showing quasi-hexagonal packing of 7–10 nm diameter intermediate filaments embedded in an osmiophilic matrix (Filshie and Rogers 1961). Many of the IF show a ring-core structure. Magnification bar 100 nm. The inset shows a cross-section of porcupine quill post-stained with $KMnO_4$ in which the IF have a ring structure (Rogers and Filshie 1962). Magnification of the inset is the same as that in the main image. (**b**) *left*, Average appearance of IF in (**a**) obtained by template matching (Fraser et al. 2003), *centre*, the (8+0) arrangement of 4-chain protofilaments, *right*, the (7+1) arrangement. (c) Vitreous section of the *stratum corneum* of human epidermis showing epidermal keratin IF with a ring-core structure. Magnification bar 50 nm. (d), Cross-section of the *stratum corneum* of neonatal mouse epidermis prepared by freeze substitution showing epidermal keratin IF with a central core. Magnification: edge of square is 10 nm (Figure (**c**) and (**d**) reprinted from Al-Amoudi et al. (2004) with permission from Wiley)
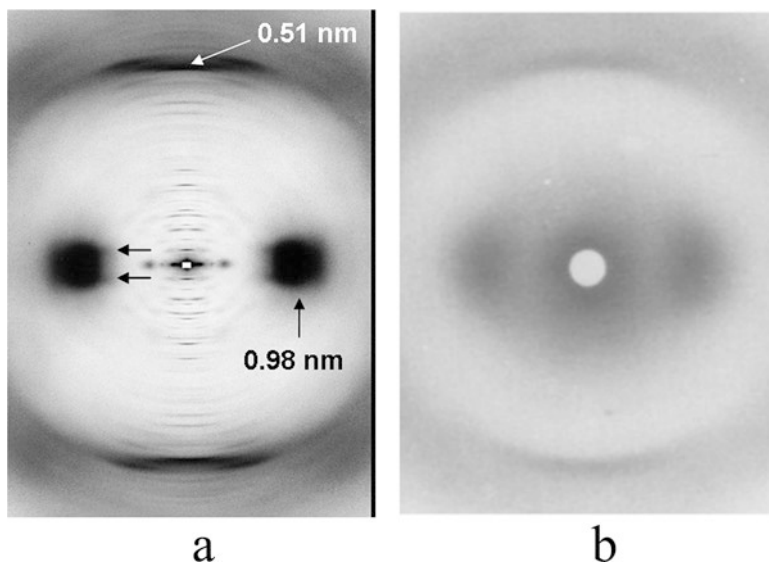
Fig. 5.2 X-ray diffraction patterns of (**a**) porcupine quill, a trichocyte keratin and (**b**) a reconstituted epidermal keratin. The former pattern is much more highly detailed and clearly shows the equatorial and near-equatorial maxima at a lateral spacing of about 0.98 nm as well as a meridional reflection at a spacing of about 0.51 nm. The axial spacing of the near-equatorial layer line is indicated by *arrows* and has a value of about 7 nm (**a** reprinted from J. Struct. Biol., 163, Parry et al., Fifty years of coiled-coils and α-helical bundles: a close relationship between sequence and structure, 36–45, 2008 with permission from Elsevier) (**b** reprinted from Steinert and Gullino (1976) with permission from Elsevier)

significant α-helical content whereas the high-sulphur fraction had none, thereby implying that the low-sulphur fraction constituted the IF and the high-sulphur fraction formed the matrix. The low-sulphur IF proteins of many trichocyte keratins have now been sequenced, and it has become clear that two families of proteins are present, termed Type I and Type II, and these are homologous not only with one another but with other members of the IF superfamily (Types III-V). Conformational predictions and crystal studies indicate that all IF molecules have a near common α-helical coiled-coil rod domain with N-terminal (head) and C-terminal (tail) domains of differing size, chemical character and sequence that are characteristic of each particular IF (see, for example, Crewther et al. 1983; Parry and Steinert 1995, 1999; Herrmann and Aebi 2004; Strnad et al. 2011). The head and tail domains in the trichocyte keratins are characteristically rich in cysteine residues.

The rod domain is made up of three segments (termed 1A, 1B and 2), and each of these has a pattern of residues having the form $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$, where positions *a* and *d* are commonly occupied by large apolar residues such as leucine, valine and isoleucine (Geisler et al. 1982; Crewther et al. 1983). The presence of a heptad repeat in an α-helix-favouring sequence is a diagnostic feature of a left-handed coiled-coil structure (Crick 1953). In the IF chains the segments 1A, 1B and 2 are
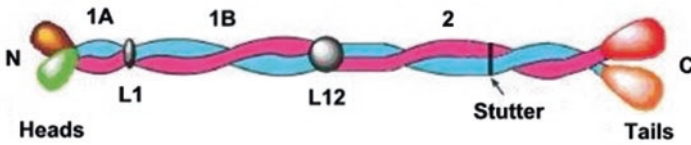
Fig. 5.3 Schematic structure of the heterodimeric (Type I/Type II) trichocyte keratin molecule. A central rod domain (1A, L1, 1B, L12 and 2) is enclosed by a head (N-terminal) and a tail (C-terminal) domain. Segments 1A, 1B and 2 have a heptad substructure thereby facilitating the formation of an α-helical left-handed coiled-coil structure. The N-terminal 30–35 residues in segment 2, however, have a hendecad substructure (closely related to the heptad repeat) that results in the α-helical strands lying approximately parallel to one another

35, 101 and 148 residues long respectively (Fig. 5.3). At the N-terminal end of segment 2 the sequence displays a hendecad (11-residue) repeat that is closely related to the heptad structure seen elsewhere (Parry 2006). The effect of this repeat is to cause the α-helical strands to lie almost parallel to the axis of the IF, rather than coil round the axis in a left-handed manner as seen elsewhere in the rod domain. Linker L1 separates segments 1A and 1B, and linker L12 separates segment 1B and 2. The linkers are generally non-α-helical. It has been shown both theoretically (Parry et al. 1977, 1985; Crewther et al. 1983) and experimentally (Steinert 1990; Hatzfeld and Weber 1990; Coulombe and Fuchs 1990) that the keratin IF molecule contains both a Type I and Type II chain, that these are parallel to one another and that they lie in axial register (Steinert et al. 1993a, b, c, 1999a, b). The overall length of the IF heterodimer (Type I/Type II) is thus about 46 nm, assuming that the axial rise per residue in a coiled-coil conformation is 0.1485 nm (Fraser et al. 1976).

Although there are no X-ray crystallographic data yet available that relate specifically to the rod domain of the trichocyte keratins the fact that the overall structure of IF molecules is expected to be the same (or very similar) across the entire family of IF proteins is very informative. Crystal data are available for virtually the entire rod domain in vimentin (Strelkov et al. 2002, 2004; Meier et al. 2009; Nicolet et al. 2010; Aziz et al. 2012; Chernyatina et al. 2012, 2015), and for parts of both lamin (Ruan et al. 2012) and epidermal keratin (Lee et al. 2012).

The structures of the high-sulphur or matrix proteins are less well understood though cysteine/proline/glycine-rich motifs five and ten residues long, postulated to form disulphide-bonded knots (Parry et al. 1979), have been found in some of these proteins, as have sequences rich in both glycine and tyrosine residues. Recent evidence has confirmed that specific interactions between matrix and IF occur *in vivo* (Fujikawa et al. 2012; Matsunaga et al. 2013) and these observations are consistent with earlier theories that IF-IF separation was determined by interactions between IF via the matrix proteins (Fraser and MacRae 1980). This was suggested from the observation that IF did not become separated from one another beyond a certain extent irrespective of the degree of swelling induced.

5.2 Axial Structure of the IF

In order to gain an understanding of the axial arrangement of the trichocyte molecules in the IF both X-ray diffraction and crosslinking studies have been employed. The former technique has focused on use of the highly oriented specimens obtainable from porcupine quill in the fully developed (oxidized) state, whereas the crosslinking methodology has the important capability of revealing the topological arrangement of the crosslinks between molecules in both the reduced and oxidized forms.

5.2.1 X-Ray Diffraction

The determination of the axial period (c) in filaments, such as those found in trichocyte keratin, depends on the identification of a series of meridional spacings (s_1, s_2, s_3, …) and choosing c such that (c/s_1, c/s_2, c/s_3, …) are, within experimental error, integral. In practice, there are two important requirements for the method to succeed. Firstly, the IF must be sufficiently well ordered to prevent near-meridional reflections arcing across the meridian, thereby leading to the possibility of recording a reflection as a meridional when, in fact, it is not. The second problem relates to having a sufficiently small X-ray beam size and sufficiently large specimen-to-film distance so that sequential orders of the period can be resolved clearly from one another. Much of the early work on the axial period in trichocyte keratin was beset by both problems and a value of 19.8 nm was initially reported (Bear 1944). This remained the generally accepted value for a quarter of a century until detailed work by Fraser and MacRae (1971), using high-resolution cameras, a very small X-ray spot size and highly oriented specimens revealed that the true period was much longer (47 nm). In a further study, Fraser et al. (1976) showed that the trichocyte IF had helical symmetry with a pitch |P| of 344.7 nm and a unit height of 47.0 nm.

5.2.2 Crosslinking

The spatial relationship between molecules in the IF can be determined by crosslinking juxtaposed lysine residues of intact IF, or sub-assemblies of them, with the periodate-cleavable bi-functional crosslinking reagent disulpho-succinimidyl-tartrate reaction (DST). This can be performed under mild conditions that do not prevent subsequent assembly of the modified proteins into IF. The crosslinked proteins are cleaved with cyanogen bromide and trypsin, and the peptides resolved by high-performance liquid chromatography (HPLC). By comparing the peptide peaks before and after crosslinking it is possible to determine which of them has been shifted. These can then be recovered for chemical characterization. Reacting these peptides with periodate reveals two peaks, each of which can be sequenced, thereby

Table 5.1 DST crosslinks for the reduced and oxidized structures of trichocyte keratin IF[a]

Reduced		
A_{11}	A_{22}	A_{12}
II 1A-05: II 2-31	II 2-31: I 2-147*	II 1A-05: I 2-147#
II 1A-31: II L12-10	II 2-36: I 2-147*	I 1B-33: II 2-76#
I 1B-33: II 1B-62	I 2-37: I 2-147*	II 1B-40: II 2-71#
II 1B-40: II IB-61	II 2-71: II 2-108*	I 1B-42:II 2-71#
II 1B-40: II 1B-62	II 2-76: II 2-98*	I 1B-71: II 2-36#
II 1B-41: II 1B-61		II L1-04:II 2-110#
I 1B-42: II 1B-61		
Oxidised		
II 1A-01: II 2-10	II 2-31: I 2-147*	II 1A-05: I 2-147#
II 1A-05: II 2-10	II 2-32: I 2-146	I 1B-33: II 2-76#
I 1A-17: II L12-10	II 2-36: I 2-147*	II 1B-40: II 2-71#
II 1A-17: II L12-10	I 2-37: I 2-147*	I 1B-42: II 2-71#
I 1B-14: I IB-71	I 2-49: I 2-132	I 1B-71: II 2-36#
I 1B-14: II 1B-71	II 2-71: II 2-108*	I 1B-101: I 2-11
I 1B-17: II 1B-66	II 2-76: II 2-98*	II L1-04: II 2-110#
II 1B-40: I 1B-42		
II 1B-41: I 1B-42		
I 1B-42: I 1B-42		
I 1B-70: II 1B-17		
I 1B-87: I L1-09		
I 1B-89: II L1-04		

Nomenclature for the crosslink pairs is chain type (I or II), segment (1A, L1, 1B, L12 or 2), and residue number within that segment. Crosslinks common to both the reduced and oxidized structures of trichocyte keratin are highlighted by an asterisk (A_{22} mode) and a hash mark (A_{12} mode). No common crosslinks are conserved for the A_{11} mode between the reduced and oxidized structures
[a]Data from Wang et al. (2000)

allowing the individual lysine residues that had been adjoined to be identified. Using all of the crosslink data collected a least-squares analysis can then be performed to ascertain the relative positions of the molecules. This approach was used successfully for epidermal keratins K1/K10 (Steinert et al. 1993a) and K5/K14 (Steinert et al. 1993b), as well as vimentin (Steinert et al. 1993c), and nestin and α-internexin (Steinert et al. 1999a, b). The trichocyte keratins were also tackled, but in this case two distinct environments were used - a reducing one to mimic the initial formation of the IF in the hair follicle, and an oxidizing one to relate to the fully developed (disulphide-bonded) form of the hair as it emerges from the scalp (Table 5.1; Wang et al. 2000; Fraser and Parry 2007). The implications were far-reaching: trichocyte keratin undergoes a profound structural rearrangement when oxidized.
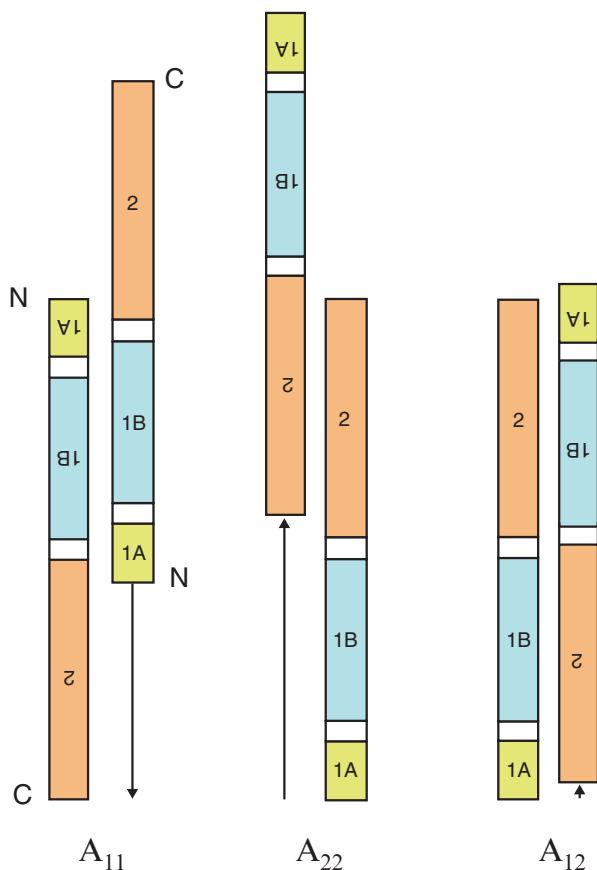
Fig. 5.4 Modes of molecular assembly initially proposed by Crewther et al. ([1983](#)). These were subsequently experimentally confirmed by Steinert et al. ([1993a](#)) and named A_{11}, A_{22} and A_{12}. These three modes are common to all classes of IF, and involve (respectively) the approximate axial alignment of antiparallel 1B segments, antiparallel 2 segments, and antiparallel rod domains. The values of A_{11}, A_{22} and A_{12} are each measured from the N-terminal residue in segment 1A in an up-pointing molecule (Figure modified from Fraser and Parry [2012](#))

In addition to the intramolecular crosslinks that confirmed that the molecule was a parallel heterodimer with its chains in axial register, the intermolecular staggers identified from the crosslinking data were easily related to three, and only three, modes of assembly termed A_{11}, A_{22} and A_{12}. A fourth mode of interaction (A_{NC}) is sometimes referred to, however, but it is effectively a combination of A_{11} and A_{22}, and is thus not unique. It is, nonetheless, directly related to the molecular overlap or gap between consecutive molecules in the molecular strands. In that respect, the use of A_{NC} has some value. The three modes are defined as follows: A_{11} is the axial stagger between antiparallel molecules that are approximately half-staggered with their 1B segments largely overlapped (Fig. [5.4a](#)), A_{22} is the axial stagger between antipar-

allel molecules that are approximately half-staggered with their 2 segments largely overlapped (Fig. 5.4b), and A_{12} is the axial stagger between antiparallel molecules that are almost completely overlapped (Fig. 5.4c). The key observations from the data are that the values of A_{12} and A_{22} are virtually unaltered between the reduced and oxidized states (about $+2\ h_{cc}$ and $+187\ h_{cc}$ respectively, where h_{cc} is the mean axial rise per residue in a coiled coil conformation), but that A_{11} changes markedly from about $-112\ h_{cc}$ (reduced) to about $-131\ h_{cc}$ (oxidized). This corresponds to an axial shift of 2.82 nm.

This molecular slippage within the IF is remarkable on three accounts. Firstly, the positions of the cysteine residues in a pair of antiparallel 1B segments are not in axial register in the reduced form and thereby incapable of forming disulphide bonds when the conditions subsequently change to an oxidizing environment. However, in the oxidized structure the molecular slippage that occurs results in the cysteine residues in the A_{11} mode being almost perfectly aligned axially (Fig. 5.5). Most of the cysteine residues would therefore be expected to form intermolecular

Fig. 5.5 Relative axial dispositions of the cysteine residues in the upward- and downward- pointing molecules in the protofilament for the (**a**) reduced and (**b**) oxidized states. The A_{11} stagger, indicated by the *arrows*, changes from −16.65 nm to −19.42 nm, thereby allowing the cysteine residues in segment 1B to lie in near axial register in the oxidized state (but not in the reduced one). Note that the cysteine residues in segment 2 lie in axial register in both the reduced and oxidized states (Reprinted from Fraser and Parry (2007) with permission from Elsevier)
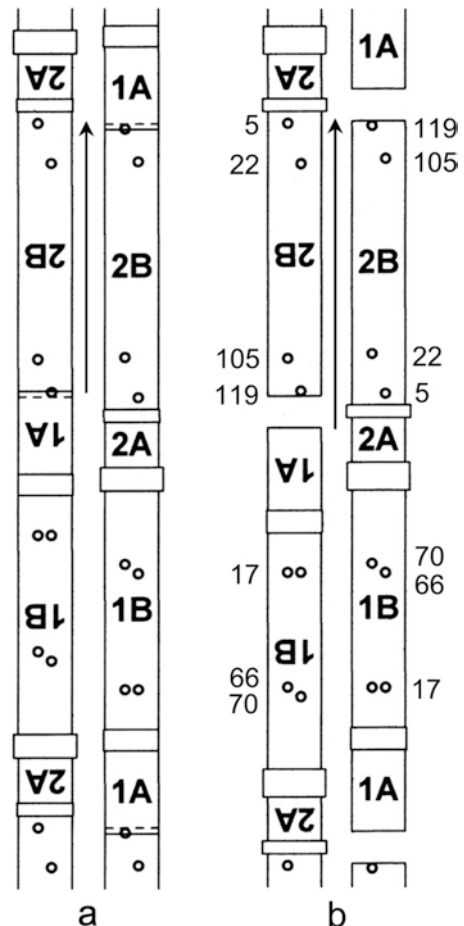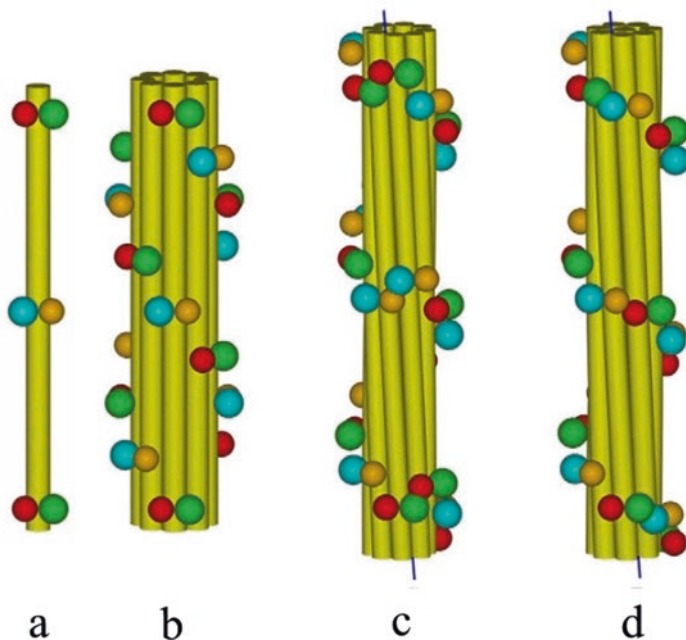
Fig. 5.6 Three-dimensional models of the structure of trichocyte keratin IF. Each IF is considered to contain eight protofilaments, and each of these consists of a pair of oppositely directed molecular strands. (**a**) A single protofilament in the reduced state, where the *yellow cylinders* represent the helical domains of the IF molecules, the *green spheres* represent the N-terminal domains of the Up molecular strands and the *blue spheres* to those of the Down molecular strands. The *red spheres* represent the C-terminal domains in the Up molecular strands and the *orange spheres* represent the C-terminal domains of the Down molecular strands. The axial period is 44.92 nm. (**b**) The reduced IF contains eight protofilaments arranged on a ring of radius 3.5 nm. Each protofilament is related to its predecessor by a rotation of 45° and an axial displacement of 16.85 nm. The terminal domains are well separated from one another and are arranged on a two-start left-handed helix, thereby giving rise at low resolution to a diagonal banding with a spacing of about 22 nm. (**c**) The (8+0) model for oxidized IF in hydrated keratin has an axial period of 47 nm, as determined from X-ray diffraction measurements. Each protofilament is related to its predecessor by a rotation of 24.3° and an axial displacement of 19.82 nm. As 47.0 nm is not a multiple of 19.82 nm there is a dislocation of 17.32 nm in the surface lattice. The terminal domains are in much closer proximity to one another than in the reduced model, and lie on a one-start helix of pitch length 23.5 nm. (**d**) As in (**c**) but for a (7+1) model. Differences from the (8+0) model are, firstly, that there are seven protofilaments in the ring surrounding a central straight one and, secondly, that the dislocation in the surface lattice is much reduced to 2.47 nm (Reprinted from Fraser and Parry (2005) with permission from Elsevier)

stabilizing disulphide bonds. Secondly, the dispositions of the heads (N-terminal domains) and tails (C-terminal domains) lying on the surface of the intermediate filament show substantial differences between the reduced and oxidized states (Fig. 5.6; Fraser and Parry 2005). In the reduced form the terminal domains are well separated from one another and are arranged on a two-start left-handed helix (Fig. 5.6b). This would give rise to a diagonal banding with a spacing of about

22 nm. In contrast, in the oxidized form the terminal domains are in much closer proximity to one another and lie on a one-start helix of pitch length 23.5 nm (Fig. 5.6c, d). Thirdly, the molecular slippage has been shown to be a simple axial translation without any associated rotation of either of the two segments involved (Fraser and Parry 2014). In the reduced model of the IF, as deduced from the cross-linking data and the position of the hydropathic stripe on each 1B chain segment, the molecules are poorly packed (near node-to-node) but in the oxidised form the molecules are in a much more compact form with close to optimal node-antinode packing. This improvement in molecular packing relates directly to the changes observed in the radial dimensions of the IF as seen between the reduced and oxidized forms (Sect. 5.4).

Since there is evidence from an energetic standpoint that the stabilization of the IF resulting from disulphide bond formation is much greater than that arising from either van der Waals interactions or hydrogen-bond formation, it seems reasonable to perform a similar least-squares analysis using the positions of the disulphide bonds as the primary data (Fraser and Parry 2012). Although these are many fewer in number than those obtained using the DST approach, the requirement for the cysteine residues to be in very near axial alignment, if disulphide bond formation is to take place, adds an important and additional constraint on the relative axial displacements of the molecules. This results in improved estimates for the values of A_{11} and A_{22} as well as a reduction in their uncertainties (Fraser and Parry 2012). Nonetheless, the values obtained did not differ greatly from those produced in the DST analyses. It is worth noting here that the appropriate alignment of the cysteine residues in the pair of antiparallel 2 segments for disulphide bond formation (A_{22}) is the same (and is unchanged) between the reduced and oxidized forms. Also, no cysteine residues are in axial positions to form disulphide bonds when the molecules are arranged in the A_{12} mode. The so-called protofilaments, which comprise a pair of antiparallel molecular strands using the A_{11} and A_{22} modes of assembly, thus contain all of the disulphide bonds present in the IF and, in this regard, the protofilament in trichocyte keratin may indeed have a true physical existence. This is in contrast to that present in other IF such as those composed of (say) vimentin, neurofilament and epidermal keratin chains, where there is little direct evidence that a protofilament is a distinct physical entity.

The formation of a disulphide bond requires a very precise spatial arrangement of the contributing cysteine residue sidechains. For this reason it is highly likely that the formation of disulphide bonds in the oxidised state will necessitate a significant local deformation of the coiled-coil structures in the immediate vicinity of the contributing cysteine residues. This will result in a reduction of the axial coherence of the coiled-coil structure, as manifested by the extent of the axial broadening of the near-equatorial layer line. It is worthy of note that the measured half-width of this layer line does indeed correspond closely to the range of lengths of coiled-coil lying between the predicted disulphide bonds (Figs. 5.5 and 5.7).

Fig. 5.7 (**a**) *x*-axis and (**b**) *z*-axis projections of the positions of the β-carbon atoms in the segment 1B dimer in the oxidized structure of trichocyte keratin IF. A cysteine residue sidechain is shown as a *yellow sphere centred* on the β-carbon atom. Observed disulphide bonds are shown in *red*. The inset in (**a**) indicates the maximum separation between β-carbon atoms involved in a disulphide bond and, as this distance is often considerably smaller than the distances marked in *red*, it follows that there must be a local disruption in the conformation of the coiled-coil to allow the disulphide bond to form, as suggested earlier by Fraser and Parry (2007) (Figure reprinted from Fraser and Parry (2014) with permission from Elsevier)
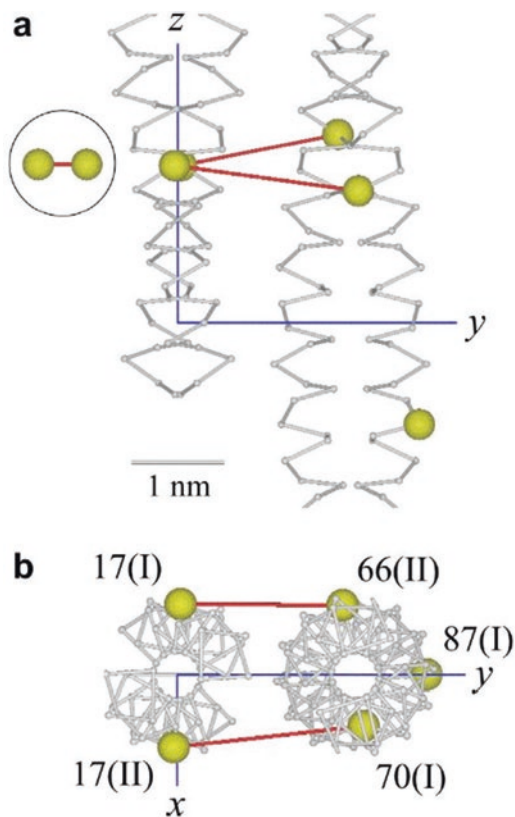
5.3 Lateral Structure of the IF

X-ray diffraction patterns obtained from highly oriented specimens of dehydrated trichocyte keratin from porcupine quill have revealed a strong but diffuse equatorial reflection with a spacing of about 0.98 nm and a near-equatorial layer line with an axial spacing of about 7 nm. The former is associated with the separation of the α-helices in the rod domain of the constituent IF molecules and the latter is an expected feature of a Type I/Type II heterodimer with a pitch length of about 14 nm. Additional low-angle equatorial reflections have been recorded (Fraser and MacRae 1958; Fraser et al. 1971) that allow the diameter of the IF (7.3 nm) to be determined from the rate of oscillation of the low-angle transform. This value is independent of the source of α-keratin used. In addition, the parameters defining the quasi-hexagonal packing of IF can be measured from this same region of the X-ray diffraction pattern. The latter vary with the origin of the hard α-keratin under observation and this is a consequence of different contents of matrix proteins in, amongst others, quill, hair and wool.

STEM data have indicated that hard-keratin IF contain about 32 chains in cross-section (Jones et al. 1997; Watts et al. 2002) and it is believed that these chains are grouped post-IF formation into eight protofilaments. Each protofilament consists of an antiparallel pair of molecular strands arranged in the A_{11} and A_{22} modes (see previous section), and each molecular strand comprises a linear array of similarly directed molecules with a small head-to-tail overlap between consecutive molecules (Fraser et al. 1990; Parry 1996; Parry et al. 2001).

5.4   Radial Projection and Surface Lattice of Trichocyte IF

Klug et al. (1958) introduced the concept of a radial projection for helical structures in which the electron density is projected radially onto a cylinder surrounding the structure. If a suitable point in the repeating unit is chosen, the radial projection of the array of points defines a surface lattice. In both cases it is convenient to plot the result in two-dimensions corresponding to an opening out of the cylinder, with the abscissa running from $-\pi$ to π as in Fig. 5.8. Klug et al. also described how the nature and the dimensions of the surface lattice could be determined from the observed X-ray diffraction pattern. In addition, they suggested that a helix could be characterized by the pitch (P) of the simplest choice of a line connecting the points in the surface lattice, a unit height (h) equal to the axial distance between successive points, and a unit twist (t) equal to the angular distance between successive points (see IUPAC-IUB Commission on Biochemical Nomenclature, 1970). For porcupine quill these parameters have the values $|P|$ = 344.7 nm, h = 47.0 nm, $|t|$ = 49.1°. The radial projection and the surface lattice are independent of the radial distribution of electron density and it is pertinent to note that evidence was presented by Fraser et al. (1965, 1967) that the rod domains of the oxidized IF molecules lay on an annulus of radius 2.9 nm (dehydrated) and 3.0 nm (hydrated). In contrast, for the reduced IF the corresponding radius was estimated to be 3.5 nm from the maximum in the radial density function obtained from freeze-dried and vitrified specimens observed in the electron microscope (Watts et al. 2002). These data suggest that a significant radial compaction occurs contemporaneously with the axial molecular re-arrangement seen between the reduced and oxidized states.

Detailed high resolution X-ray diffraction data from the meridional and near-meridional regions of fully keratinized (oxidized) trichocyte keratin IF have been obtained by Fraser, MacRae and colleagues (Fraser and MacRae 1983, 1985; Fraser et al. 1976, 1986, 1990). From these data they have described two possible surface lattice structures – the seven- and the eight-protofilament models - in terms of the axial projections of the a and b surface lattice vectors (z_a and z_b respectively) and the changes in unit twist involved (t_a and t_b respectively) (Fig. 5.7). Note that there are no data available to distinguish between the surface lattices shown and their mirror images about a vertical line. The axial repeat h (47 nm), z_a (7.42 nm) and z_b (19.79 nm) are identical for each model. The physical significance of z_a and z_b is that the former is the axial distance between successive surface lattice points, and the latter

Fig. 5.8 Fraser and MacRae (1983, 1985) have described two possible surface lattice structures based on the idea of fractured helices, and these are referred to as the seven- and eight-protofilament models. They do not differ in terms of the axial repeat h (47 nm), or the projections of the a and b surface lattice vectors (z_a=7.42 nm and z_b=19.79 nm respectively). They do differ, however, in the changes in azimuth involved (t_a and t_b respectively). There are no data currently available that distinguish between these surface lattices and their mirror images about a vertical line. The values of t_a and t_b for the seven-protofilament model shown here are $-4\pi/7$ and $+2\pi/7$ respectively. This model, with an additional eighth protofilament centrally located, is referred to as the (7+1) model and, on the basis of the currently available data, is a likely structure for the oxidized trichocyte keratin IF. The precise nature of the fracture is unknown but the original suggestion of a helical dislocation along the broken line is perhaps the simplest
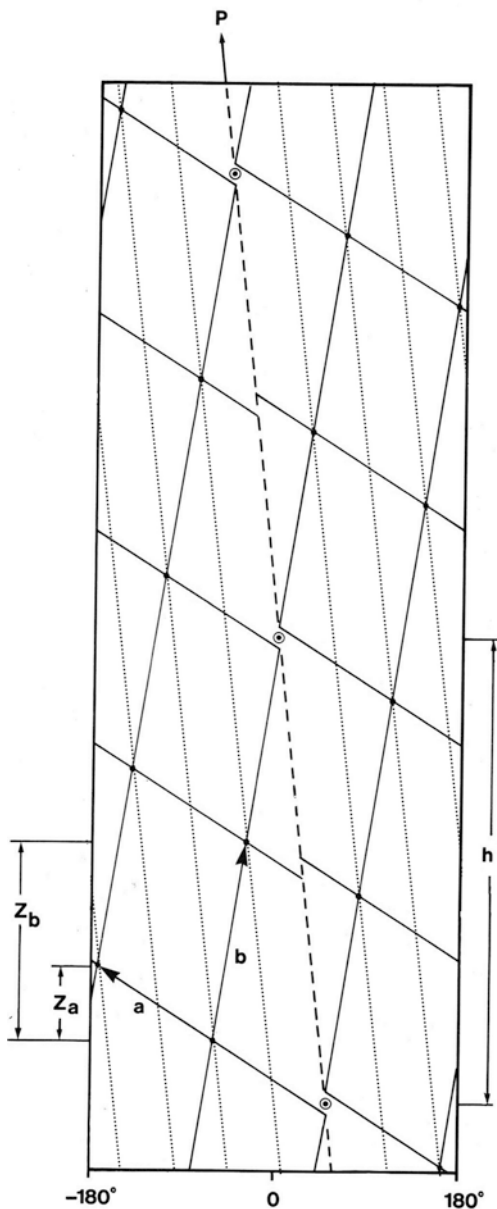
is the axial stagger between adjacent protofilaments, each of which would have four chains in section. It is of numerological interest that z_b (19.79 nm) corresponds almost exactly to seven times the change in A_{11} that occurs between the oxidized and reduced states (7 × 2.82 nm = 19.74 nm), though the structural implication of this observation has yet to be fully appreciated. The values of t_a and t_b for the

eight-protofilament structure are $-\pi/2$ and $+\pi/4$ respectively and those for the seven-protofilament model are $-4\pi/7$ and $+2\pi/7$ respectively. The former consists of a ring of eight protofilaments arranged on a constant radius. The second model consists of seven protofilaments arranged on a constant but slightly smaller radius than for the eight-protofilament model. Although, in the latter case, the concept that an eighth protofilament might be centrally located was not considered at that time by Fraser and MacRae it is now common practice to refer to this as the (7+1) model. It is important to note that the X-ray data on which this surface lattice structure is based pertains specifically to that collected for the oxidized IF.

Electron microscope data obtained by Watts et al. (2002) on decorated reduced filaments, which were analysed by Fraser et al. (2003), provide direct support for an eight-protofilament model for the structure of the reduced IF. This is referred to as the (8+0) model and has the lattice dimensions as follows: axial repeat = 44.92 nm, $z_a = 11.23$ nm, $t_a = -\pi/2$, $z_b = 16.85$ nm, $t_b = \pi/4$. These dimensions are necessarily quite different than those for the two possible models for the oxidized structure: the (7+1) model with an axial repeat $c = 47.0$ nm, $z_a = 7.42$ nm, $t_a = -4\pi/7$, $z_b = 19.79$ nm, $t_b = +2\pi/7$, and the (8+0) structure with an axial repeat $c = 47.0$ nm, $z_a = 7.42$ nm, $t_a = -\pi/2$, $z_b = 19.79$ nm, $t_b = +\pi/4$ (Fig. 5.8).

Early TEM data indicated a ring-core structure (now considered as possibly relating to a (7+1) structure) for the oxidized IF (Filshie and Rogers 1961; Millward 1970) though some observations yielded a ring structure (possibly consistent with an (8+0) structure). As noted earlier these differences in appearance in the TEM were a source of controversy over the years as to which set of observations represented the *in vivo* structure. Later, cryo-electron microscopy on IF produced at the very earliest stage of development (the reduced state) suggested a ring structure i.e. no core (Watts et al. 2002). The interpretation of the TEM observations have remained unclear, and while they may indeed reflect different stages in the development of the hair fibre there remains the distinct possibility that differences in the staining protocols used have resulted in observations that are not capable of a unique interpretation. Currently, the two models for the structure of the oxidized IF cannot be clearly differentiated on the basis of the TEM observations, the X-ray diffraction data or the physical dimensions of the IF (see Fraser et al. 1986 and Fraser and Parry 2003), and both models remain viable alternatives.

In summary, in axial projection the IF in the reduced state would contain eight protofilaments arranged regularly on an annulus of radius 3.5 nm (hydrated). In the oxidized state the structure of the IF would consist of either (a) seven protofilaments lying on an annulus of electron density (radius 3.0 nm), hence leading to the observed decrease in radius between the reduced and oxidized cases, and an eighth protofilament, which is straight and located at the centre of the IF or (b) eight protofilaments lying on a ring of radius 3.0 nm, but more closely packed than in the reduced model. In the reduced and oxidized models the protofilaments lying within the annulus are unlikely to remain strictly parallel to the axis of the IF and are likely to exhibit modest supercoiling. The concept that there might be different structures for the reduced and oxidized forms of the IF was not envisaged at the time that the observations were first made.

5.5   Summary

Fifty years ago it would have been hard to imagine that trichocyte (hair) keratin IF would have two different structures, depending on its stage of keratinization. With hindsight, of course, that possibility might have been envisaged much earlier had the results of the histology and cytochemistry research been better understood and appreciated. On the basis of the currently available data it appears that the structure of the filaments in the lower part of the follicle, in the presence of a reducing environment, differs from that of the filaments in the upper part of the follicle where the cells are densely packed with IF and are exposed to an oxidizing environment.

The observation that molecular slippage allowed many disulphide bonds to be formed within the IF neatly explains why the hair fibre is so resilient and tough, and thus so well suited to its role as both a thermo-regulator and protector of the animal against its external environment.

It is of interest that sections of epidermal IF (Fig. 5.1d, e) also exhibit a ring-core appearance in the electron microscope similar to that seen in trichocyte IF. This could indicate that epidermal keratin molecules, although containing very few cysteine residues in their rod domains and with markedly different head and tail domain characteristics from those in the trichocyte keratins, might also undergo a similar compaction in a transition from a reducing to an oxidizing environment. More data will be required, however, before this concept can be considered proven.

References

Al-Amoudi A, Chang J-J, Leforestier A, McDowall A, Salomin LM, Norlén LPO, Richter K, Sartori Blanc N, Studer D, Dubochet J (2004) Cryo-electron microscopy of vitreous sections. EMBO J 23:3583–3588

Astbury WT, Woods HJ (1930) The X-ray interpretation of the structure and elastic properties of hair keratin. Nature 126:913–914

Astbury WT, Woods HJ (1933) X-ray studies on the structure of hair, wool and related fibres II. The molecular structure and elastic properties of hair keratin. Phil Trans Roy Soc Lond A 232:333–394

Aziz A, Hess JF, Budamagunta MS, Voss JC, Kuzin AP, Huang YJ, Xiao R, Montelione GT, FitzGerald PG, Hunt JF (2012) The structure of vimentin linker 1 and rod 1b domains characterized by site-directed spin-labeling electron paramagnetic resonance (SDSL-EPR) and X-ray crystallography. J Biol Chem 287:28349–28361

Bear RS (1944) X-ray diffraction studies on protein fibers. II Feather keratin, porcupine quill and clam muscle. J Am Chem Soc 66:2043–2050

Birbeck MSC, Mercer EH (1957) The electron microscopy of human hair follicle. J Biophys Biochem Cytol 3:203–214

Chernyatina AA, Nicolet S, Aebi U, Herrmann H, Strelkov SV (2012) Atomic structure of the vimentin central α-helical domain and its implications for intermediate filament assembly. Proc Natl Acad Sci U S A 109:13620–13625

Chernyatina AA, Guzenko D, Strelkov SV (2015) Intermediate filament structure: the bottom-up approach. Curr Opin Cell Biol 32:65–72

Crewther WG, Fraser RDB, Lennox FG, Lindley H (1965) The chemistry of keratins. Adv Protein Chem 20:191–346

Crewther WG, Dowling LM, Steinert PM, Parry DAD (1983) Structure of intermediate filaments. Int J Biol Macromol 5:267–274

Crick FHC (1953) The packing of alpha-helices – simple coiled-coils. Acta Cryst 6:689–697

Coulombe PA, Fuchs E (1990) Elucidating the early stages of keratin filament assembly. J Cell Biol 111:153–169

Filshie BK, Rogers GE (1961) The fine structure of α-keratin. J Mol Biol 3:784–786

Fraser RDB, MacRae TP (1958) Structural implications of the equatorial X-ray diffraction pattern of α-keratin. Biochim Biophys Acta 29:229–240

Fraser RDB, MacRae TP (1971) Structure of α-keratin. Nature (London) 233:138–140

Fraser RDB, MacRae TP (1980) Molecular structure and mechanical properties of keratins. In: Vincent JFV, Currey JD (eds) The mechanical properties of biological materials. Cambridge University Press, Cambridge, pp 211–246

Fraser RDB, MacRae TP (1983) The structure of the α-keratin microfibril. Biosci Rep 3:517–525

Fraser RDB, MacRae TP (1985) Intermediate filament structure. Biosci Rep 5:573–579

Fraser RDB, Parry DAD (2003) Macrofibril assembly in trichocyte (hard α-) keratins. J Struct Biol 142:319–325

Fraser RDB, Parry DAD (2005) The three-dimensional structure of trichocyte (hard α-) keratin intermediate filaments: features of the molecular packing deduced from the sites of induced crosslinks. J Struct Biol 151:171–181

Fraser RDB, Parry DAD (2007) Structural changes in the trichocyte intermediate filaments accompanying the transition from the reduced to the oxidized form. J Struct Biol 159:36–45

Fraser RDB, Parry DAD (2012) The role of disulfide bond formation in the structural transition observed in the intermediate filaments of the developing hair. J Struct Biol 180:117–124

Fraser RDB, Parry DAD (2014) Keratin intermediate filaments: differences in the sequences of the type I and type II chains explain the origin of the stability of an enzyme-resistant four-chain fragment. J Struct Biol 185:317–326

Fraser RDB, MacRae TP, Miller A (1964) The coiled-coil model of α-keratin structure. J Mol Biol 10:147–156

Fraser RDB, MacRae TP, Miller A (1965) X-ray diffraction pattern of α-fibrous proteins. J Mol Biol 14:432–442

Fraser RDB, MacRae TP, Parry DAD (1967) The substructure of the α-keratin microfibril. In: Crewther WG (ed) Symposium on fibrous proteins (Australia 1967), Butterworths, Australia, pp 279–286

Fraser RDB, MacRae TP, Millward GR, Parry DAD, Suzuki E, Tulloch PA (1971) The molecular structure of keratins. Appl Polym Symp No 18:65–83

Fraser RDB, MacRae TP, Rogers GE (1972) Keratins – their composition, structure and biosynthesis. Thomas, Springfield

Fraser RDB, MacRae TP, Suzuki E (1976) Structure of the α-keratin microfibril. J Mol Biol 108:435–452

Fraser RDB, MacRae TP, Parry DAD, Suzuki E (1986) Intermediate filaments in α-keratins. Proc Natl Acad Sci U S A 83:1179–1183

Fraser RDB, MacRae TP, Parry DAD (1990) The three-dimensional structure of IF. In: RD Goldman, PM Steinert (eds) Cellular and molecular biology of intermediate filaments. Plenum, New York, pp 205–231

Fraser RDB, Steinert PM, Parry DAD (2003) Structural changes in trichocyte keratin intermediate filaments during keratinization. J Struct Biol 142:266–271

Fujikawa H, Fujimoto A, Farooq M, Ito M, Shimomura Y (2012) Characterization of the human hair keratin-associated protein 2 (KRTAP2) gene family. J Invest Dermatol 132:1806–1813

Geisler N, Kaufman E, Weber K (1982) Protein-chemical characterization of three structurally distinct domains along the protofilament unit of 10 nm filaments. Cell 30:277–286

Hatzfeld, M. and Weber, K. 1990. The coiled-coil of in vitro assembled keratin filaments is a heterodimer of type I and II keratin: use of site-specific mutagenesis and recombinant protein expression. J Cell Biol 110:1199–1210

Herrmann H, Aebi U (2004) Intermediate filaments: Molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. Annu Rev Biochem 73:749–789

Jones LN (1976) Studies on microfibrils from α-keratin. Biochim Biophys Acta 446:515–524

Jones LN, Simon M, Watts NR, Booey FP, Steven AC, Parry DAD (1997) Intermediate filament structure: Hard alpha-keratin. Biophys Chem 68:83–93

Klug A, Crick FHC, Wyckoff HW (1958) Diffraction by helical structures. Acta Cryst 11:199–213

Lee CH, Kim MS, Chung BM, Leahy DJ, Coulombe PA (2012) Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. Nat Struct Mol Biol 19:707–715

MacArthur I (1943) Structure of α-keratin. Nature 152:38–41

Matsunaga R, Abe R, Ishii D, Watanabe S-I, Kiyoshi M, Nöcker B, Tsuchiya M, Tsumoto K (2013) Bidirectional binding property of high glycine-tyrosine keratin-associated protein contributes to the mechanical strength and shape of hair. J Struct Biol 183:484–494

Meier M, Padilla GP, Herrmann H, Wedig T, Hergt M, Patel TR, Stetefeld J, Aebi U, Burkhard P (2009) Vimentin coil 1A – a molecular switch involved in the initiation of filament elongation. J Mol Biol 390:245–261

Millward GR (1970) The sub-structure of α-keratin microfibrils. J Ultrastruct Res 31:349–355

Nicolet S, Herrmann H, Aebi U, Strelkov SV (2010) Atomic structure of vimentin coil 2. J Struct Biol 170:369–376

Parry DAD (1996) Hard α-keratin intermediate filaments: an alternative explanation of the low-angle equatorial X-ray diffraction pattern, and the axial disposition of putative disulphide bonds in the intra- and inter-protofilamentous networks. Int J Biol Macromol 19:45–50

Parry DAD (2006) Hendecad repeat in segment 2A and linker L2 of intermediate filament chains implies the possibility of a right-handed coiled coil structure. J Struct Biol 155:370–374

Parry DAD, Steinert PM (1995) Intermediate filament structure, Springer, Heidelberg, pp 1–183

Parry DAD, Steinert PM (1999) Intermediate filaments: molecular architecture, assembly, dynamics and polymorphism. Q Rev Biophys 32:99–187

Parry DAD, Crewther WG, Fraser RDB, MacRae TP (1977) Structure of α-keratin: structural implication of the amino acid sequences of the type I and type II chain segments. J Mol Biol 113:449–454

Parry DAD, Fraser RDB, MacRae TP (1979) Repeating patterns of amino acid residues in the sequences of some high-sulphur proteins from α-keratin. Int J Biol Macromol 1:17–22

Parry DAD, Steven AC, Steinert PM (1985) The coiled-coil molecules of intermediate filaments consist of two parallel chains in exact axial register. Biochem Biophys Res Commun 127:1012–1018

Parry DAD, Marekov LN, Steinert PM (2001) Subfilamentous protofibril structures in fibrous proteins: cross-linking evidence for protofibrils in intermediate filaments. J Biol Chem 276:39253–39258

Rogers GE (1959) Electron microscope studies of hair and wool. Ann N Y Acad Sci 83:378–399

Rogers GE, Filshie BK (1962) Electron staining and fine structure of keratins. Proc. Intern. Congr. Electron Microscopy, 5th Philadelphia, vol 2, O-2. Academic, New York

Ruan J, Xu C, Bian C, Lam R, Wang JP, Kania J, Min J, Zang J (2012) Crystal structures of the coil 2B fragment and the globular tail domain of human lamin B1. FEBS Lett 586:314–318

Steinert PM (1990) The two-chain coiled-coil molecule of native epidermal keratin intermediate filaments is a type I –type II heterodimer. J Biol Chem 265:8766–8774

Steinert PM, Gullino MI (1976) Bovine epidermal keratin filament assembly *in vitro*. Biochem Biophys Res Commun 70:221–227

Steinert PM, Marekov LN, Fraser RDB, Parry DAD (1993a) Keratin intermediate filament structure: crosslinking studies yield quantitative information on molecular dimensions and mechanism of assembly. J Mol Biol 230:436–452

Steinert PM, Marekov LN, Parry DAD (1993b) Conservation of the structure of keratin intermediate filaments: molecular mechanism by which different keratin molecules integrate into preexisting keratin intermediate filaments during differentiation. Biochemistry 32:10046–10056

Steinert PM, Marekov LN, Parry DAD (1993c) Diversity of intermediate filament structure: evidence that the alignment of coiled-coil molecules in vimentin is different from that in keratin intermediate filaments. J Biol Chem 268:24916–24925

Steinert PM, Chou Y-H, Prahlad V, Parry DAD, Marekov LN, Wu KC, Jang S-I, Goldman RD (1999a) A high molecular weight intermediate filament-associated protein in BHK-21 cells is nestin, a type VI intermediate filament protein: limited co-assembly in vitro to form heteropolymers with type III vimentin and type IV α-internexin. J Biol Chem 274:9881–9890

Steinert PM, Marekov LN, Parry DAD (1999b) Molecular parameters of type IV α-internexin and type IV-type III α-internexin-vimentin copolymer intermediate filaments. J Biol Chem 274:1657–1666

Strelkov SV, Herrmann H, Geisler N, Wedig T, Zimbelmann R, Aebi U (2002) Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. EMBO J 21:1255–1266

Strelkov SV, Schumacher J, Burkhard P, Aebi U, Herrmann H (2004) Crystal structure of the human lamin A coil 2B dimer: implications for the head-to-tail association of nuclear lamins. J Mol Biol 343:1067–1080

Strnad P, Usachov V, Debes C, Gräter F, Parry DAD, Omary MB (2011) Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins. J Cell Sci 124:4221–4232

Wang H, Parry DAD, Jones LN, Idler WW, Marekov LN, Steinert PM (2000) In vitro assembly and structure of trichocyte keratin intermediate filaments: a novel role for stabilization by disulfide bonding. J Cell Biol 151:1459–1468

Watts NR, Jones LN, Cheng N, Wall JS, Parry DAD, Steven AC (2002) Cryo-electron microscopy of trichocyte (hard α-keratin) intermediate filaments reveals a low-density core. J Struct Biol 137:109–118

Chapter 6
Crystallographic Studies of Intermediate Filament Proteins

Dmytro Guzenko, Anastasia A. Chernyatina, and Sergei V. Strelkov

Contents

Abstract Intermediate filaments (IFs), together with microtubules and actin microfilaments, are the three main cytoskeletal components in metazoan cells. IFs are formed by a distinct protein family, which is made up of 70 members in humans. Most IF proteins are tissue- or organelle-specific, which includes lamins, the IF proteins of the nucleus. The building block of IFs is an elongated dimer, which consists of a central α-helical 'rod' domain flanked by flexible N- and C-terminal domains. The conserved rod domain is the 'signature feature' of the IF family. Bioinformatics analysis reveals that the rod domain of all IF proteins contains three α-helical segments of largely conserved length, interconnected by linkers. Moreover, there is a conserved pattern of hydrophobic repeats within each segment, which includes heptads and hendecads. This defines the presence of both left-handed and almost parallel coiled-coil regions along the rod length. Using X-ray crystallography on multiple overlapping fragments of IF proteins, the atomic structure of the nearly complete rod domain has been determined. Here, we discuss some specific challenges of this procedure, such as crystallization and diffraction data phasing by molecular replacement. Further insights into the structure of the coiled coil and the terminal domains have been obtained using electron paramagnetic resonance measurements on the full-length protein, with spin labels attached at specific positions. This atomic resolution information, as well as further interesting findings, such as

D. Guzenko • A.A. Chernyatina • S.V. Strelkov (✉)
Laboratory for Biocrystallography, Department of Pharmaceutical
and Pharmacological Sciences, Katholieke Universiteit Leuven, Leuven, Belgium
e-mail: sergei.strelkov@kuleuven.be

the variation of the coiled-coil stability along the rod length, provide clues towards interpreting the data on IF assembly, collected by a range of methods. However, a full description of this process at the molecular level is not yet at hand.

Abbreviations

EPR	Electron Paramagnetic Resonance
IF	Intermediate Filament
MF	Microfilament
MR	Molecular Replacement
MT	Microtubule
RMSD	Root-Mean-Square Deviation
SDSL	Site-Directed Spin Labelling
ULF	Unit-Length Filament

6.1 Introduction

Intermediate filaments (IFs), together with microtubules (MTs) and microfilaments (MFs), are the three principal systems of cytoskeletal filaments found in metazoan cells (Herrmann et al. 2007). While the two other filament systems are assembled from globular subunits, the building unit of IFs is a rod-like dimer based on a coiled-coil structure. The human genome contains about 70 different IF protein genes (Szeverenyi et al. 2008). Most of those are differentially expressed in various cell and tissue types, including tissues of epithelial, mesodermal and to a lesser extent endodermal origin (Steinert and Roop 1988).

IF protein sequences reveal a common tripartite organization, consisting of a readily predicted α-helical central 'rod' domain, which is flanked by non-helical amino- ('head') and carboxy terminal- ('tail') domains. Traditionally, the IF proteins are classified into six types. The two most abundant types are acidic keratins (type I) and basic keratins (type II). Keratins are found in epithelial cells, where IFs play an essential role in cell integrity and tissue resilience (Pekny and Lane 2007). A distinctive feature of keratins is their preferential assembly into heterodimers, composed of one acidic and one basic chain, while most other IF proteins are found as homodimers. Type III embraces proteins found in different tissues with varying functions, including vimentin. Historically, this has been used in many structural studies. Type IV contains neurofilaments expressed in axons of neuronal cells. Type V encompasses the lamins, the IF proteins of the cell nucleus. Lamins form the nuclear lamina, a filamentous network that supports the nuclear envelope on its
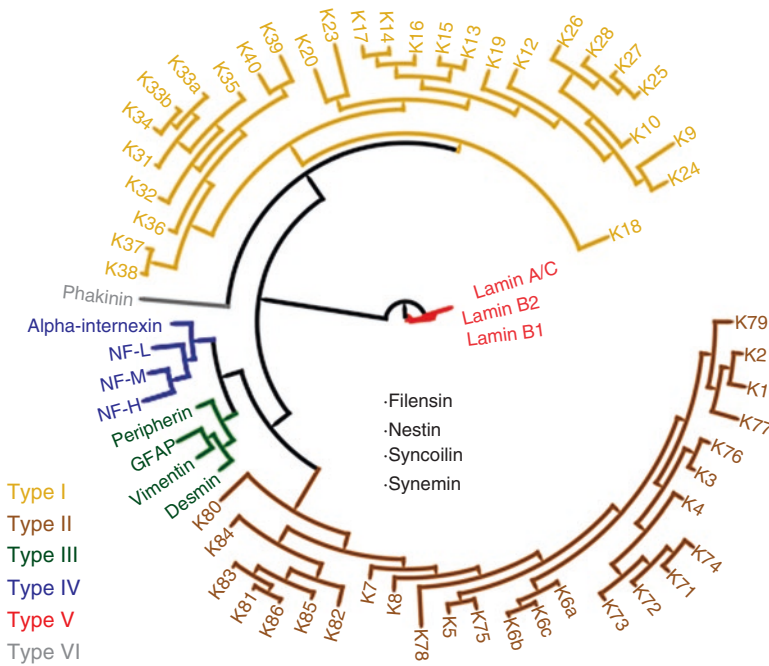
Fig. 6.1 IF protein classification presented as a phylogenetic tree. The tree was built using PhyML (Guindon et al. 2010) based on a multiple sequence alignment of the central rod domains. Very distant IF proteins could not be reliably placed on the tree. They are labelled as Type VI, although their classification in the literature may vary

inner side. These proteins can be seen as evolutionary ancestors of cytoplasmic IF proteins (Peter and Stick 2015). A majority of IF proteins can be reliably placed in a homology tree based on similarity of their rod domains, including cytoplasmic IF proteins (types I to IV) and nuclear lamins (type V). Finally, type VI contains 'orphan' IF proteins, which are more distant in terms of their amino acid sequences (Fig. 6.1).

Cytoplasmic IF proteins assemble into 10–12 nm wide filaments which lack polarity. This is in contrast to the polar MTs and MFs (Robert et al. 2016). The assembly process does not require any additional proteins or energy source and can be reproduced *in vitro* (Herrmann et al. 2004). The assembly of cytoplasmic proteins starts upon the lateral association of two dimers in an antiparallel, half-staggered way (the so-called A_{11} mode), yielding a tetramer which can be found as a soluble species in low molarity neutral buffers (Mucke et al. 2004) (Fig. 6.2). An increase of the ionic strength and an adjustment of pH induces the further lateral association of tetramers, yielding structures known as the 'unit-length filaments' (ULF), which contain about eight tetramers in the case of human vimentin (Mucke et al. 2004). Finally, the ULFs anneal longitudinally to yield long filaments, which undergo further compaction, apparent from a decrease of their diameter. In cyto-plasmic IF proteins, the presence of the head domain is essential for proper filament
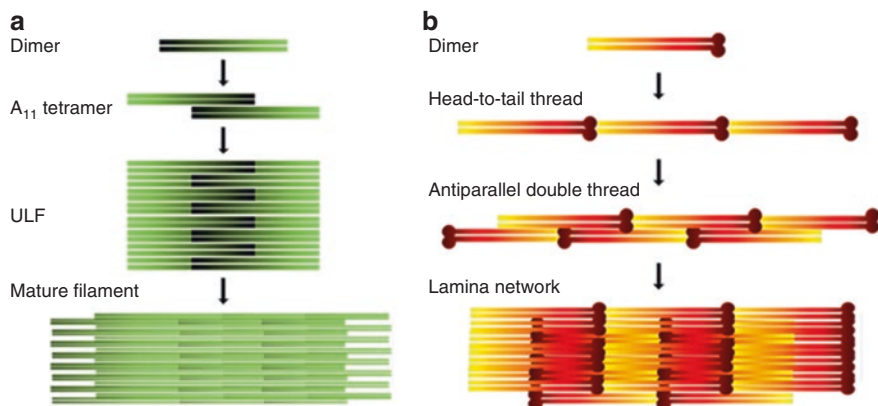
Fig. 6.2 Schematic representation of IF assembly. (**a**) Assembly of cytoplasmic IF proteins typically starts from heterodimers (keratins) or homodimers (other proteins), advancing to antiparallel half-staggered tetramers, unit-length filaments (ULFs) and ultimately mature filaments. (**b**) The nuclear lamins have a distinctly different pathway of assembly, as the dimers tend to interact longitudinally in a head-to-tail fashion. *Dark circles* designate the globular tail domains

assembly (Herrmann et al. 1996). The role of the tail domain, however, is more obscure. Some IF proteins, *e.g.* phakinin, lack the tail domain altogether, while its truncation in IF proteins of types I to III leads to largely normal filament formation, albeit with relaxed width control (Herrmann et al. 1996). Interestingly, nuclear lamins have a clearly different assembly pathway than the cytoplasmic IFs. At an early stage, the lamin dimers appear to assemble longitudinally in a head-to-tail fashion. These longer threads also associate in an antiparallel way, ultimately producing the intertwined, heterogeneous filaments forming the nuclear lamina (Herrmann et al. 2007) (Fig. 6.2).

Due to the primary mechanical function of IFs in both the cytoplasm and the nucleus, they are critically involved in shaping cell organelles, cell migration and division (Goldman et al. 1996). In addition, by interacting with other proteins and DNA, IFs participate in the up- and down-regulation of many cellular processes, including the stress response (Ivaska et al. 2007). IFs also interact with heat shock proteins and are major substrates for stress kinases (Toivola et al. 2010), while phosphorylation is the most common regulatory mechanism of IF assembly (Snider and Omary 2014). Like other stress proteins, IFs are capable of inhibiting apoptosis, which correlates with their demonstrated involvement in various types of cancers (Satelli and Li 2011).

Despite several decades of research, our current understanding of the IF structure and assembly pathway remains limited, which contrasts with the much more detailed structural description that has been achieved for both MTs and MFs. In particular, direct 3D visualization of both cytoplasmic and nuclear IFs using (cryo) electron microscopy has provided only partial insights, which may be linked to a limited degree of order in the assembled IF structures (Goldie et al. 2007). In this regard, a systematic 'bottom-up' approach, starting with the detailed analysis of the

elementary IF dimer structure, appears necessary. Indeed, knowledge of the elementary dimer at atomic resolution provides a solid basis for using the available data from other methods towards building convincing models of the assembly process and the mature filament architecture. Here we will briefly review recent achievements towards dissecting the three-dimensional structure of the IF dimer using X-ray crystallography and related techniques, and explain the remaining questions and experimental challenges. For further reading, please refer to our recent review paper in *Current Opinion in Cell Biology* (Chernyatina et al. 2015) which contains a more detailed description of the IF dimer structure *per se*, and a chapter in *Methods in Enzymology* (Chernyatina et al. 2016), which focuses on the experimental and *in silico* methods currently used to analyse IF proteins.

6.2   Primary Structure of IF Proteins

Amino acid sequence analysis of IF proteins can be a vast source of information on the IF structure, especially considering the detailed understanding of the coiled coil motif available today. As an 'all-in-one' reference on IF genes, proteins and mutations, one can conveniently consult the Human Intermediate Filament Database at http://interfil.org (Szeverenyi et al. 2008).

The homologous α-helical rod domain is a key characteristic of all IF proteins. The length of the rod domain is close to 310 residues in cytoplasmic IF proteins and 350 residues in lamins. The α-helical structure is interrupted in two places by short sequences called the linkers (L1 and L12), resulting in three segments named coil1A, coil1B and coil2. While the exact definition of the ends of the α-helical segments varies slightly depending on the particular secondary structure prediction algorithm, the length of coil1A seems to be close to 42 residues in nearly all IF proteins, coil1B is ~102 residues in cytoplasmic IFs and ~140 residues in lamins, and coil2 is ~140 residues in nearly all IF proteins (Fig. 6.3). The sequence conservation across different IF proteins is especially high for most of coil1A and also for the region at the end of coil2. These two so-called 'consensus motifs', each about 30 amino acids long, appear to be important for filament assembly (Herrmann and Aebi 2004) (Fig. 6.4).

Most of the rod domain reveals a heptad repeat pattern, characteristic of a classical left-handed coiled coil. This motif can be readily detected by standard tools such as COILS (Lupas et al. 1991). In addition, a few years back we proposed a more nuanced approach towards detecting the exact coiled-coil structure within the rod domain (Chernyatina et al. 2012). This approach utilizes the prediction of solvent exposure of residues along each sequence using, for example, output from the NetSurfP algorithm (Petersen et al. 2009). Indeed, the residues in the core positions of a coiled coil (usually occupied by Leu, Ala, Glu, Val, Ile and Lys (Parry et al. 2008)) are less exposed than the residues on its outer side. The task is therefore to scan the protein sequence for repeat patterns in the occurrence of the most buried residues.
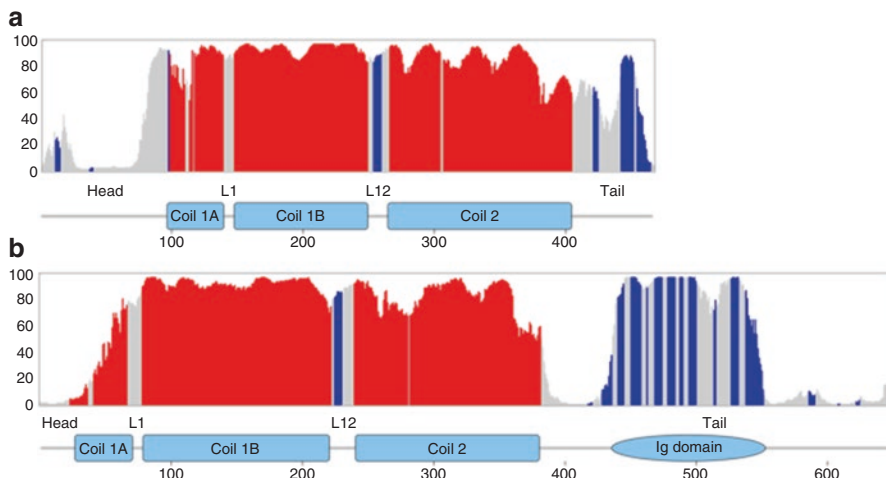
Fig. 6.3 Overall structure of the IF monomer, as revealed by bioinformatic analysis of the amino acid sequences of human vimentin (**a**) and human lamin A (**b**). The height of the graph shows the degree of order predicted using DISOPRED3 (Jones and Cozzetto 2015), plotted versus the residue number. The most probable secondary structure predicted using Jpred (Drozdetskiy et al. 2015) is additionally indicated by the colour: α-helices in *red*, β-strands in *blue*, random coil in *grey*. The domain structure of the dimer is shown below as a graph. Rectangles show the continuous α-helical segments of the rod domain, and an oval corresponds to the β-structural Ig fold within the lamin tail

This way it becomes possible to detect other types of repeat patterns that could be present in coiled coils beyond the classical heptads (Gruber and Lupas 2003).

Such an approach, applied to the α-helical region of an IF dimer, works surprisingly well. Most importantly, it shows that the beginning of the coil2 α-helical segment contains several instances of the hendecad pattern *abcdefghijk*, with the hydrophobic core formed by the positions *a*, *d/e* and *h*. This pattern is predicted to form a right-handed coiled coil with a very large pitch, *i.e.* practically a parallel bundle (Kuhnel et al. 2004). We suggest that such a structure should be better classified as a *coiled coil* rather than a generic α-helical bundle, since this structure features a specific interaction of the core-forming residues. The hendecad pattern near the N-terminus of coil2 is consistently predicted for all IF proteins (Fig. 6.4), in line with earlier analyses (Parry 2006). In addition, there are single instances of a hendecad near the end of coil1B and also in the middle of coil2 (Fig. 6.4). A single hendecad is nothing different than a heptad with a four-residue insertion, a situation known as a 'stutter' in a regular left-handed coiled coil (Brown et al. 1996). Importantly, for IF types I through IV, *i.e.* for a vast majority of IF proteins, the hydrophobic patterns within each of the three α-helical segments (*i.e.* specific sequences of heptads and hendecads) are absolutely conserved, as apparent from sequence alignments (Fig. 6.4). This repeat pattern is also found in nuclear lamins, save the fact that their coil1B contains six additional heptads (42 residues). Only some orphan IF proteins from type VI partially deviate from this principle.
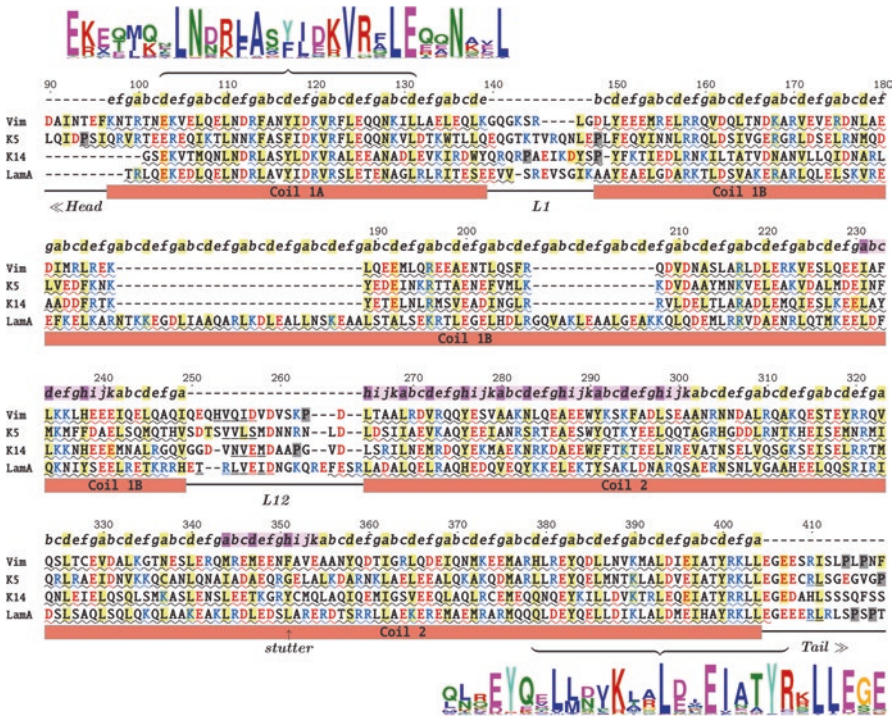
Fig. 6.4 Primary structure of the IF rod domain. Amino acid sequences for human vimentin (numbered above the sequence), keratins 5 and 14, and lamin A are aligned. Secondary structure prediction using the Jpred algorithm (Drozdetskiy et al. 2015) is shown for each sequence (wave underline indicates α-helix, simple underline indicates β-strand). Residues predicted by the NetSurfP algorithm (Petersen et al. 2009) to be buried in the hydrophobic core are highlighted in *yellow*. Heptad and hendecad assignments are indicated, with the latter also highlighted in *violet*. The hendecad repeat in the middle of coil2 is traditionally labelled as 'stutter'. Proline residues are highlighted in *grey*. Coil1A and coil2 consensus motifs are indicated with braces and presented by sequence logos. The motifs were defined upon application of the MEME motif elicitation algorithm (Bailey et al. 2006) to all human IF sequences (Szeverenyi et al. 2008) (Figure adapted from (Chernyatina et al. 2015) with permission)

We therefore suggest that the 'sequence signature' of the IF protein family is defined by several features, namely:

1. The presence of a central α-helical domain
2. Division of the α-helical domain into three segments of mostly conserved length
3. Conservation of the hydrophobic pattern, which includes hendecads at specific positions and heptads elsewhere.

Although not 'carved in stone', these three signature features (as opposed to solely the presence of a long heptad-containing region) should, in our opinion, guide an assignment of a given protein to the IF family. In the past, several proteins, including crescentin (Ausmees et al. 2003) and isomin (Mencarelli et al. 2011), were proposed to represent a new IF family member, or at least be dubbed 'IF-like'.

Moreover, it is easy to see that any two coiled-coil-forming sequences would have some degree of sequence homology, due to the prevalence of hydrophobic and polar residues in the core and exposed positions, respectively. Correcting for this intrinsic feature is important in preventing any premature conclusions on 'IF-likeliness'. Moreover, while filament formation is a key functional property of the IF family *per se*, this is not a decisive feature. For example, totally unrelated designed coiled-coil-forming peptides were shown to self-assemble into filaments. However, these filaments lacked such intrinsic characteristics of IFs as width control, and exhibited very different mechanical properties, including much lower flexibility and extensibility (MacPhee and Woolfson 2004).

The properties of the N- and C-terminal sequences of IF proteins are very different from that of the α-helical rod. A key overall feature of these domains is a low degree of predicted order (Fig. 6.3), with the most prominent exception being the globular immunoglobulin fold present within the lamin tail. A parallel observation is a large variability in the sequences of terminal regions within one IF type, and even a complete lack of sequence conservation between many types. The varied terminal regions appear to contribute significantly to the differences in the biochemical properties of IF proteins, their assembly pathways and, ultimately, to specific functions in the cell.

6.3 Experimental Studies of the Elementary Dimer

Despite relatively good structural predictions being possible, based on the amino acid sequences of IF proteins, experimental verification of the elementary dimer structure had a major impact towards substantiating our understanding of the IFs. Here, X-ray crystallography has played a central role (Chernyatina et al. 2015; Chernyatina et al. 2016). A number of challenges had to be overcome along the way. Intact full-length IF proteins are inherently not suitable for crystallization due to their ability to self-assemble into filaments. After removal of the flexible N- and C-terminal domains, which are typically required for the correct assembly process (Mucke et al. 2004), an isolated rod domain should, in principle, remain soluble and thus be appropriate for crystallization. However, crystals of the highly-elongated, over 300 residues-long, entire rod domain, have not been reported for any IF protein thus far. Instead, a 'divide-and-conquer' approach, based on preparation, crystallization and X-ray analysis of shorter overlapping segments, has been the most productive (Strelkov et al. 2001). Indeed, thanks to the 'linear nature' of the coiled-coil rod, the folding and structure of individual fragments is likely to be representative of a full-length protein. This is in stark contrast with globular proteins, where fragmentation of a domain typically leads to serious structural consequences and should be avoided.

Table 6.1 lists all fragments of the IF rod that have been resolved using X-ray crystallography to date. X-ray structure determination of these fragments was based upon producing them using recombinant *E. coli* expression in milligram amounts,

Table 6.1 Crystal structures of coiled-coil fragments of human vimentin (VIM), keratin (K) and lamin A (LA) and B1 (LB1)

PDB code	Resid.	Structural part and reference	Resol. (Å)	Phasing method	Dimers/asu
1GK7*	VIM 102–138	Monomeric coil1A, does not form a coiled coil (Strelkov et al. 2002)	1.4	MR	Monomer
3G1E	VIM 101–139	Coil1A with a stabilizing mutation Y117L (Meier et al. 2009)	1.83	MR	1
3SSU	VIM 99–189	Fragment including coil1A, L1 and part of coil1B; N-terminus up to residue 143 is disordered (Chernyatina et al. 2012)	2.6	SM	1
3S4R	VIM 99–189	Fragment including coil1A, L1 and part of coil1B, with a stabilizing mutation Y117L (Chernyatina et al. 2012); Coil1A part forms an antiparallel assembly of 4 helices	2.45	SM	1
3UF1	VIM 144–251	Coil1B (Aziz et al. 2012); two dimers form a biological tetramer	2.81	SM	2
3SWK	VIM 153–238	Major part of vimentin coil1B (Chernyatina et al. 2012)	1.7	SM	1
4YV3*	VIM 161–238	Major part of vimentin coil1B forming a non-natural trimer (Chernyatina et al. 2016)	2.0	SM	trimer
4YPC*	VIM 161–243	Major part of vimentin coil1B forming a non-natural trimer (Chernyatina et al. 2016)	1.44	MR	1/3 trimer
3KLT*	VIM 263–334	First half of coil2 (Nicolet et al. 2010). N-termini of dimers form a non-natural 4-helix antiparallel bundle	2.7	HA (SmCl$_3$)	2
3TRT	VIM 261–335	First half of coil2 with a dimerizing mutation L265C; Disulfide bond C265–C265 is resolved (Chernyatina and Strelkov 2012)	2.3	SM	1
1GK4	VIM 328–411	Second half of coil2 (Strelkov et al. 2002)	2.3	HA (Sm acetate)	3
1GK6	VIM 385–412	C-terminal section of coil2 fused to GCN4 zipper (Strelkov et al. 2002)	1.9	MR	1/2
3TNU	K14/K5 295–422 350–477	Second half of coil2 of keratin 5 and keratin 14 heterodimer (Lee et al. 2012)	3.01	SM	2

Table 6.1 (continued)

PDB code	Resid.	Structural part and reference	Resol. (Å)	Phasing method	Dimers/asu
1X8Y	LA: 305–387	Second half of lamin A coil2 (Strelkov et al. 2004)	2.2	MR	1/2
3V5B	LA: 313–386	Second half of lamin A coil2 (Bollati et al. 2012)	3.00	MR	1/2
3V4Q	LA: 313–386	Second half of lamin A coil2 (R335W mutant) (Bollati et al. 2012)	3.06	MR	1/2
3V4W	LA: 313–386	Second half of lamin A coil2 (E347K mutant) (Bollati et al. 2012)	3.7	MR	1/2
3TYY	LB1: 311–388	Second half of lamin B1 coil2 (Ruan et al. 2012); Disulphide bond between Cys317 residues (d position)	2.4	MR	1
2XV5	LA: 328–398	Second half of lamin A coil2 with a small tail section (Kapinos et al. 2011); biological significance of the observed antiparallel right-handed coiled-coil dimer is not clear	2.4	SM	1

Structures deviating from the biologically relevant dimers are marked with *asterisks*. The phasing methods used are abbreviated as follows: *MR* molecular replacement, *SM* selenomethionine phasing, *HA* phasing using soaked heavy atoms

followed by purification and screening for crystallizability in a trial-and-error fashion. Detailed descriptions of these procedures may be found in Chernyatina et al. (2016). As for other proteins, crystallizability of such fragments depends on multiple properties and, as such, is difficult to predict. In our hands, the success rate of crystallization does not exceed ~15 % of all IF fragments prepared.

In addition, crystallographic studies of IF rod fragments have been complemented by data for the full-length protein, obtained using a very different experimental technique, namely electron paramagnetic resonance with site-directed spin labels (SDSL-EPR). In a series of studies, such labels were systematically attached in various positions, covering practically the whole length of the human vimentin sequence (Hess et al. 2004 ; Chernyatina et al. 2016). Thereafter, the EPR measurements were used to reveal the proximity of the labels at various IF assembly stages. Within a homodimeric rod, proximity of a label to its counterpart in another chain is characteristic for a core position of the coiled coil. In addition, the EPR data can reveal the degree of structural order in the vicinity of the labelled residue. As will be explained below, the SDSL-EPR measurements have provided important complementarity to the fragment-based crystallographic data on the rod domain, and have also shed light on the conformation of the head and tail domains.

6.4 Crystallographic Challenges of Coiled-Coil Fragments

Experience has shown that shorter, more compact coiled-coil fragments have a higher chance of crystallization, with the longest IF protein fragment crystallized thus far being 106 residues in length (Chernyatina et al. 2015). At the same time, shorter fragments are more likely to assemble in an aberrant way. The latter may include failure to oligomerize into a coiled coil, the formation of coiled-coil oligomers other than a dimer, and even antiparallel and staggered arrangements. A few examples of such assembly difficulties are a vimentin fragment corresponding to coil1A that failed to form a coiled coil (PDB code 1GK7, (Strelkov et al. 2002)), as well as the N-terminal portion of vimentin coil2, which formed an aberrant antiparallel tetramer (PDB code 3KLT, (Nicolet et al. 2010)). Another interesting case relates to the structures obtained for multiple vimentin fragments belonging to coil1B. Here, two longer fragments practically corresponding to the entire segment, namely residues 144–251 (PDB code 3UF1, (Aziz et al. 2012)) and 153–238 (PDB code 3SWK, (Chernyatina et al. 2012)), have formed proper coiled-coil dimers, while two other fragments that are slightly shorter at the N-terminal side, which included residues 161–238 (PDB code 4YV3) and 161–243 (PDB code 4YPC) respectively, were found in an aberrant trimeric state (Chernyatina et al. 2016) (Table 6.1).

Keeping in mind these difficulties, several approaches have been used to secure proper parallel dimer formation. In particular, for a vimentin fragment 385–412 fusion with the GCN4 leucine zipper was used (Strelkov et al. 2002). The introduction of a cysteine residue in a core position near the N-terminal end of the fragment was used for the beginning portion of vimentin coil2, whereby a formed disulphide bond induced the correct overall conformation (PDB code 3TRT, (Chernyatina and Strelkov 2012)). Finally, the Y117L mutation has been successfully used to stabilize the coil1A region of vimentin, resulting in a dimeric coiled coil (PDB code 3G1E, (Meier et al. 2009)). In addition, recent developments in structural bioinformatics, such as improved sequence-based predictions of the coiled-coil oligomeric state (Vincent et al. 2013) and *in silico* assessment of the effect of mutations on coiled-coil stability (Potapov et al. 2015) may further help to improve the IF fragment design for future crystallographic studies.

A further, well-known difficulty of crystallographic studies is the fact that diffraction phases cannot be measured directly (the so-called 'phase problem', (Taylor 2010)). Experimental means to obtain the missing phasing information, either by using SeMet-labelled protein for crystallization or by soaking the native crystals in heavy atom solutions, have been used for the majority of IF fragment structures determined to date (Chernyatina et al. 2016) (Table 6.1). However, these experimental procedures are not always straightforward, especially when reproducibility of the crystals is a problem. An alternative approach is the phasing of diffraction data through a purely computational procedure known as 'molecular replacement' (MR), which involves placing the known 3D structure of a homologous protein (the 'search model') into the new crystal lattice (Rossmann 1990). In fact, phasing by MR may
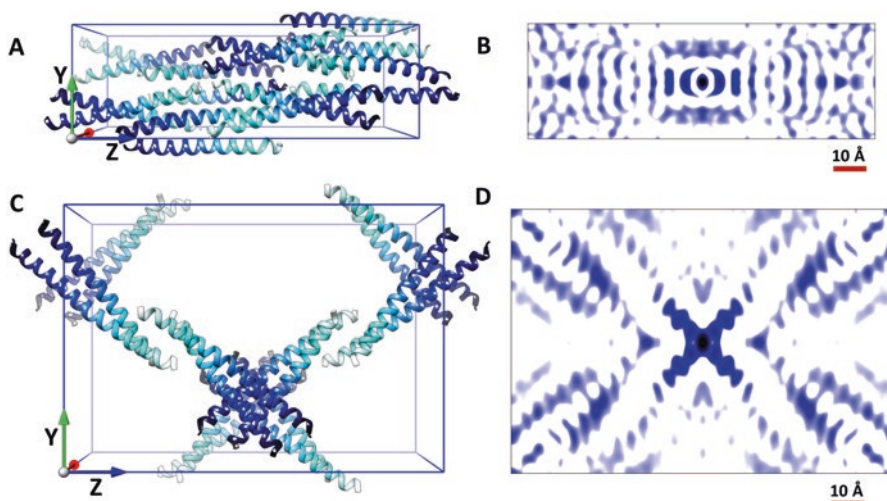
Fig. 6.5 (**a, c**) Crystallographic packing arrangements for two vimentin fragment structures, PDB codes 3G1E (**a**) and 3SSU (**c**). The fragments are shown as ribbon diagrams, coloured in gradient from N-terminus (*light blue*) to C-terminus (*dark blue*). (**b, d**) Cross-sections (YZ plane) of the native Patterson function maps for the two fragments, with darker colours corresponding to higher values. The Patterson map is shifted by Y/2 and Z/2 for illustrative purposes

appear the most straightforward possibility for IF fragments, since the coiled-coil geometry is quite simple, and one could expect that finding a suitable search model should not be a problem. Indeed, most coiled coils can be idealised within 1 Å root-mean-square deviation (RMSD) of atomic positions using compact parameterization with a rather narrow distribution of the parameters (Grigoryan and DeGrado 2011). This, in principle, allows the researcher to quickly generate a limited set of models, such as that using the CCBuilder online tool (Wood et al. 2014)). At least one of the generated models is likely to be sufficiently close to the target structure, which should, in principle, enable a successful application of the MR procedure. Recently, several MR pipelines designed specifically for coiled-coil proteins have been introduced, which are based on modelling the structure with the Rosetta suite (Kaufmann et al. 2010; Rämisch et al. 2015). Moreover, we have found that analysis of the Patterson function calculated with the native protein diffraction data can often give an idea about the packing of the crystal lattice. The Patterson function is equal to a Fourier summation, including diffraction intensities as coefficients (thus no phase information is required), and shows all interatomic distances in the crystal lattice (Patterson 1935). Specifically, we have found that the orientation of the coiled-coil dimer is often apparent from a line of equally spaced (5.43 Å) Patterson peaks next to the origin (Fig. 6.5). Such a periodicity emerges as a product of the number of residues per α-helical turn (3.62) and the translation along the coiled-coil axis per residue. The latter figure for a heptad-based coiled coil is 1.50 Å on average (based on calculations employing the Twister algorithm (Strelkov and Burkhard 2002)), *i.e.* slightly less than the corresponding translation along the axis of an

isolated α-helix (1.56 Å). This analysis of the Patterson function is of assistance in determining the possible orientation of the coiled coil axis (provided it remains straight or almost straight in the crystals, as is indeed seen for nearly all IF rod fragments studied) and may help to reduce the dimensionality of the MR search.

In fact, successful phasing by MR has been reported for a number of IF fragment structures (Table 6.1). However, despite all of the above considerations, our experience indicates that MR can never be guaranteed to work successfully. Our main conclusion is that these difficulties are not due to the lack of accuracy of the search model *per se* (although for longer coiled coils, conformational variability can indeed complicate things further), but rather by the fundamental challenges of the MR procedure when the search model is a coiled coil. The complication here is the intrinsic symmetry of the latter: in particular, a heptad-based structure overlaps onto itself upon a seven-residue register shift. In this case, there may be numerous ways of positioning the model in the crystal lattice in terms of both rotation and translation, and each could produce reasonable agreement with the experimental data (as judged by the likelihood criteria utilized in modern MR algorithms). As a result, the contrast of the correct MR solution might be quite low. In fact, we have observed that in certain extreme cases (apparently dependent on a particular space group and packing arrangement) the MR procedure fails in an *a posteriori* test using a 'perfect' structure, *i.e.* the one previously obtained through experimental phasing. In summary, experience with IF fragments indicates that the application of MR phasing to coiled coils remain a challenge.

6.5 Elementary Dimer Structure

A large number of IF protein fragments have now been solved at atomic resolution. These collectively cover the whole length of the rod domain, with the exception of the linker L12 region. Most of these fragments have come from human vimentin, which has served as a 'model' IF protein in multiple studies, while some X-ray data have also been obtained for lamin and keratin fragments (Fig. 6.6 and Table 6.1). Importantly, superposition of several overlapping fragments typically results in a good agreement between the structures (Fig. 6.6c, d). This provides a positive evaluation of the approach. In addition, fragment superpositions give an idea about the accuracy of the atomic structure determination, as indeed the relatively flexible coiled-coil fragments may be conformationally-distorted, depending on the particular crystal contacts that they engage in. The overall model of the full-length cytoplasmic IF dimer is shown in Fig. 6.6e. This model also reflects the expected cloud of conformations of the N- and C-terminal domains, which remain largely disordered, at least at early IF assembly stages.

Figure 6.7 shows detailed snapshots of the two key repetitive structural motifs of the rod domain: the left-handed heptad-based coiled coil, as exemplified by the N-terminal section of vimentin coil1B, and the parallel α-helical bundle based on
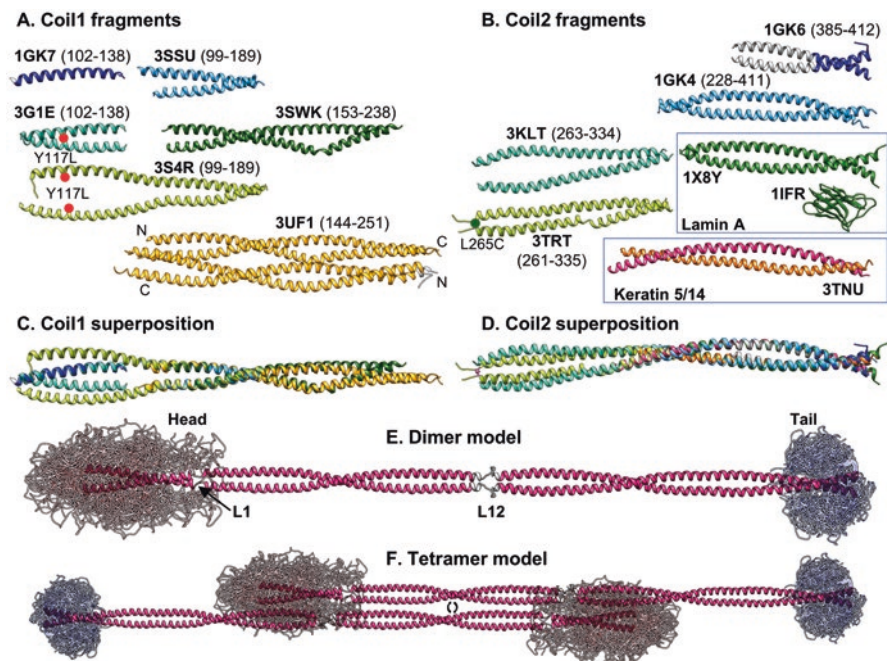
Fig. 6.6 Molecular structure of the IF dimer. (**a**) Crystal structures of all coil1 fragments available to date, shown as ribbon diagrams. The fragments, all from human vimentin, are labelled by their PDB codes, with the residue range indicated in brackets (Aziz et al. 2012; Chernyatina et al. 2012; Meier et al. 2009; Strelkov et al. 2002). The fragments are oriented in the same way. The artificial mutation Y117L is marked with a red dot. (**b**) Crystal structures of coil2 fragments of vimentin, the keratin 5/14 heterodimer and lamin A. The structure of the immunoglobulin fold present in the tail domain of lamin A (PDB code 1IFR) is also shown (Dhe-Paganon et al. 2002). The 1GK6 structure contains a leucine zipper domain (*grey*) fused to the vimentin fragment. Artificial disulphide bridge in the 3TRT fragment is marked with a *green* dot. The available structures of the coil2 and tail fragments of lamin B1 (Ruan et al. 2012) are highly similar to the corresponding fragments of lamin A, and are not shown. (**c**) Least-squares superposition of all available coil1 fragments. (**d**) Least-squares superposition of coil2 fragments. (**e**) Atomic model of a complete vimentin dimer. The model, assuming perfect twofold symmetry, was produced by computational modelling using all available crystal structures of fragments as templates. The regions of the rod for which no experimental structures are available and which were modelled *ab initio* are shown in *grey*. The poorly ordered head and tail domains (*semi-transparent red and blue* respectively) are best given by a cloud-like representation (40 modelled conformations). (**f**) Atomic model of the vimentin tetramer. The alignment of two dimers was based on the 3UF1 structure. The twofold axis of the tetramer is indicated (Figure reprinted, with permission, from Chernyatina et al. (2015))

hendecad repeats, as exemplified by the N-terminal section of coil2. As may be seen, the actual involvement of residues in the hydrophobic core in either case (positions *a* and *d* of heptads and positions *a, d, e* and *h* of hendecads) correlates very well with the sequence-based prediction of the least exposed residues. Even more importantly, excellent correlation has been observed between the pattern of core positions found in the crystal structures and those detected using the SDSL-EPR
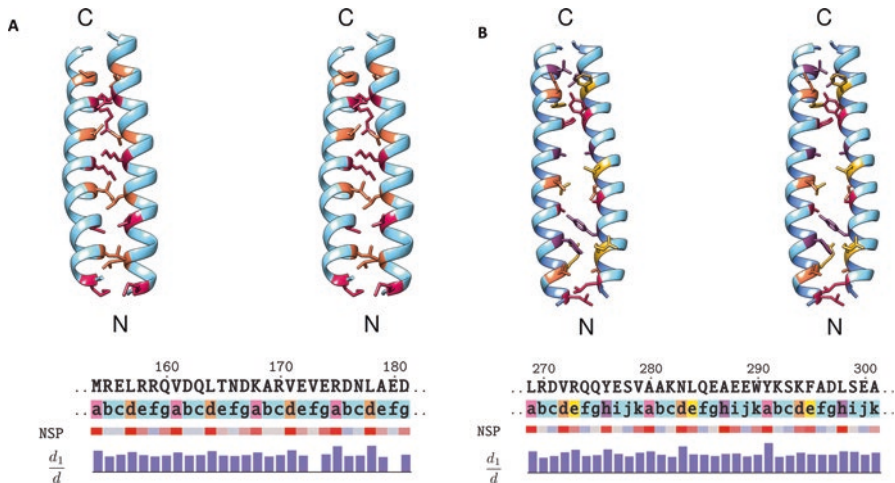
Fig. 6.7 Stereo view of the two types of coiled-coil structure found in IF dimers. (**a**) Classical left-handed coiled coil formed by the N-terminal part of coil1B (residues 154–181 corresponding to three heptads, PDB code 3SSU, (Chernyatina et al. 2012)). (**b**) N-terminal part of coil2 forming a parallel α-helical bundle (residues 269–301 corresponding to three hendecads). The crystal structure (PDB code 3TRT) had some irregularities apparently linked to the introduction of an artificial disulphide bridge at the N-terminus and crystal packing (Chernyatina and Strelkov 2012). Shown here is a re-modelled structure obtained using Rosetta fold-and-dock protocol (Kaufmann et al. 2010) with soft distance constraints derived from the crystal structure. The residues are coloured according to the position in the heptad or hendecad repeat. Amino-acid sequences are given below the structures, with heptad and hendecad repeats highlighted. Sequence-based prediction of solvent exposure using NetSurfP (Petersen et al. 2009) is shown as a colour scale (*red* – buried in the hydrophobic core, *blue* – exposed to solvent). The d_1/d values calculated from the SDSL EPR experiments (Aziz et al. 2012; Hess et al. 2006) are shown as a bar graph. The d_1/d values correlate with the proximity of the residues

approach for the full-length protein (Aziz et al. 2012; Chernyatina et al. 2016). This correlation is also very apparent for the representative structures shown in Fig. 6.7.

Another important discovery that emerged during the studies of multiple IF rod fragments was that the coiled-coil stability varies greatly along the rod length. In particular, the fragment corresponding to vimentin coil1A (and thus containing the complete N-terminal consensus motif of the rod) formed a coiled coil of only marginal stability, since its melting temperature was only 32°C (Meier et al. 2009). A single engineered point mutation (Y117L) was capable of increasing the melting temperature up to 78 °C. Likewise, a fragment including the hendecad repeats from the N-terminal part of coil2 was found to be monomeric in solution (Nicolet et al. 2010). Analysis of the corresponding parallel bundle structure revealed that its hydrophobic core contains several alanine and tyrosine residues as well as one asparagine, all of which are not optimal for stability. Indeed, the loose packing of the coiled-coil core in this region is quite apparent (Fig. 6.7b). Importantly, there are some indications that variations in coiled-coil stability along the rod length are needed if correct IF assembly is to proceed. For instance, introduction of the Y117L

mutation into the full-length vimentin resulted in a greatly reduced capacity for filament elongation (Meier et al. 2009). We hypothesize that the assembly may require the unzipping of some parts of the rod, but more research is necessary to test this idea.

Not surprisingly, crystallization of IF fragments containing poorly ordered parts has proved to be a challenge. Despite our extensive efforts, no human vimentin or lamin A fragments containing the linker L12 region, even in the presence of adjacent coiled-coil, has ever been crystallized. Regarding the flexible N- and C-terminal domains, two short overlapping fragments of the vimentin head domain (residues 59–71 and 66–78) have been crystallized recently when bound to a human leukocyte antigen (PDB codes 4MD0, 4MDJ, (Scally et al. 2013)). At the same time, some information on the terminal domains has been gathered using SDSL-EPR as well as chemical cross-linking (Herrmann and Aebi 2004), allowing the creation of initial models indicating their overall topology (Aziz et al. 2010; Hess et al. 2013). In particular, within an individual vimentin dimer, the head domains appear to fold back onto one side of the coil1 domain, breaking the overall symmetry of the dimer. This asymmetry may help to differentiate between possible inter-dimeric interactions and thus contribute to defining a specific pathway of filament assembly.

6.6 Conclusions and Outlook

The research outlined above has provided a nearly complete structure of the IF rod at atomic resolution. Structurally, the key elements of the IF dimer are the two longer coiled-coil segments, the relatively rigid coil1B and coil2, connected by the presumably flexible linker L12. In addition, there is a shorter and seemingly dynamic coil1A region, as well as the flexible N- and C-terminal domains.

A further important direction is to obtain atomic detail on the assembly stage beyond the dimer, i.e. the tetramer. Recently, a crystal structure of the vimentin coil1B fragment (PDB code 3UF1, (Aziz et al. 2012)) has revealed the first tetrameric assembly of this kind, consisting of two dimers running in opposite directions. It should be noted that molecular contacts seen in this structure are compatible with the data obtained for the full-length tetramer by both chemical crosslinking (Steinert et al. 1993) and SDSL-EPR (Hess et al. 2004). This has made 3D modelling of the full-length vimentin tetramer possible (Fig. 6.6f), although more high-resolution experimental data on the tetramer, including for other IF types, are clearly needed.

Ultimately, atomic resolution data should help us to understand the IF assembly process, which has been characterized by a range of both *in vivo* and *in vitro* experiments. Slowly but surely, such aspects as the specific inter-dimer interactions driving the filament assembly, the particular pathway and kinetics of assembly, and the external factors affecting it, are becoming clearer. Nonetheless, a complete mechanistic model that would cover both the course and the result of the assembly *i.e.* the mature IF, is still a thing of the future.

The atomic resolution data obtained for 'model' proteins such as vimentin should be used, *via* a detailed sequence-based bioinformatics analysis, towards a better structural understanding of the whole IF family. However, care should be taken here in order to avoid excessive extrapolations. Although keratins and also other IF proteins of one or even several different types are sometimes seen co-assembling into 'mixed' filaments (Wickert et al. 2005 ; Steinert et al. 1999), there can also be drastic structural differences across them in other cases. For instance, despite a largely similar elementary dimer structure, the assembly pathway of nuclear lamins is radically different from that of cytoplasmic IFs (Fig. 6.2), and the resulting lamin filaments seem to lack width control (Dittmer and Misteli 2011).

References

Ausmees N, Kuhn JR, Jacobs-Wagner C (2003) The bacterial cytoskeleton: an intermediate filament-like function in cell shape. Cell 115(6):705–713

Aziz A, Hess JF, Budamagunta MS, Voss JC, Fitzgerald PG (2010) Site-directed spin labeling and electron paramagnetic resonance determination of vimentin head domain structure. J Biol Chem 285(20):15278–15285

Aziz A, Hess JF, Budamagunta MS, Voss JC, Kuzin AP, Huang YJ, Xiao R, Montelione GT, FitzGerald PG, Hunt JF (2012) The structure of vimentin linker 1 and rod 1B domains characterized by site-directed spin-labeling electron paramagnetic resonance (SDSL-EPR) and X-ray crystallography. J Biol Chem 287(34):28349–28361

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34(suppl 2):W369–W373

Bollati M, Barbiroli A, Favalli V, Arbustini E, Charron P, Bolognesi M (2012) Structures of the lamin A/C R335W and E347K mutants: implications for dilated cardiolaminopathies. Biochem Biophys Res Commun 418(2):217–221

Brown JH, Cohen C, Parry DAD (1996) Heptad breaks in alpha-helical coiled coils: stutters and stammers. Proteins 26(2):134–145

Chernyatina AA, Strelkov SV (2012) Stabilization of vimentin coil2 fragment via an engineered disulfide. J Struct Biol 177(1):46–53

Chernyatina AA, Nicolet S, Aebi U, Herrmann H, Strelkov SV (2012) Atomic structure of the vimentin central alpha-helical domain and its implications for intermediate filament assembly. Proc Natl Acad Sci U S A 109(34):13620–13625

Chernyatina AA, Guzenko D, Strelkov SV (2015) Intermediate filament structure: the bottom-up approach. Curr Opin Cell Biol 32C:65–72

Chernyatina AA, Hess JF, Guzenko D, Voss JC, Strelkov SV (2016) How to study intermediate filaments in atomic detail. Methods Enzymol 568:3–33

Dhe-Paganon S, Werner ED, Chi YI, Shoelson SE (2002) Structure of the globular tail of nuclear lamin. J Biol Chem 277(20):17381–17384

Dittmer TA, Misteli T (2011) The lamin protein family. Genome Biol 12(5):222

Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure prediction server. Nucleic Acids Res 43(W1):W389–W394

Goldie KN, Wedig T, Mitra AK, Aebi U, Herrmann H, Hoenger A (2007) Dissecting the 3-D structure of vimentin intermediate filaments by cryo-electron tomography. J Struct Biol 158(3):378–385

Goldman RD, Khuon S, Chou YH, Opal P, Steinert PM (1996) The function of intermediate filaments in cell shape and cytoskeletal integrity. J Cell Biol 134(4):971–983

Grigoryan G, DeGrado WF (2011) Probing designability via a generalized model of helical bundle geometry. J Mol Biol 405(4):1079–1100

Gruber M, Lupas AN (2003) Historical review: another 50th anniversary–new periodicities in coiled coils. Trends Biochem Sci 28(12):679–685

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307–321

Herrmann H, Haner M, Brettel M, Muller SA, Goldie KN, Fedtke B, Lustig A, Franke WW, Aebi U (1996) Structure and assembly properties of the intermediate filament protein vimentin: the role of its head, rod and tail domains. J Mol Biol 264(5):933–953

Herrmann H, Aebi U (2004) Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular Scaffolds. Annu Rev Biochem 73:749–789

Herrmann H, Kreplak L, Aebi U (2004) Isolation, characterization, and in vitro assembly of intermediate filaments. Methods Cell Biol 78:3–24

Herrmann H, Bar H, Kreplak L, Strelkov SV, Aebi U (2007) Intermediate filaments: from cell architecture to nanomechanics. Nat Rev Mol Cell Biol 8(7):562–573

Hess JF, Budamagunta MS, Voss JC, FitzGerald PG (2004) Structural characterization of human vimentin rod 1 and the sequencing of assembly steps in intermediate filament formation in vitro using site-directed spin labeling and electron paramagnetic resonance. J Biol Chem 279(43):44841–44846

Hess JF, Budamagunta MS, Shipman RL, FitzGerald PG, Voss JC (2006) Characterization of the linker 2 region in human vimentin using site-directed spin labeling and electron paramagnetic resonance. Biochemistry 45(39):11737–11743

Hess JF, Budamagunta MS, Aziz A, FitzGerald PG, Voss JC (2013) Electron paramagnetic resonance analysis of the vimentin tail domain reveals points of order in a largely disordered region and conformational adaptation upon filament assembly. Protein Sci 22(1):47–55

Ivaska J, Pallari HM, Nevo J, Eriksson JE (2007) Novel functions of vimentin in cell adhesion, migration, and signaling. Exp Cell Res 313(10):2050–2062

Jones DT, Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31(6):857–863

Kapinos LE, Burkhard P, Herrmann H, Aebi U, Strelkov SV (2011) Simultaneous formation of right- and left-handed anti-parallel coiled-coil interfaces by a coil2 fragment of human lamin A. J Mol Biol 408(1):135–146

Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J (2010) Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry 49(14):2987–2998

Kuhnel K, Jarchau T, Wolf E, Schlichting I, Walter U, Wittinghofer A, Strelkov SV (2004) The VASP tetramerization domain is a right-handed coiled coil based on a 15-residue repeat. Proc Natl Acad Sci U S A 101(49):17027–17032

Lee CH, Kim MS, Chung BM, Leahy DJ, Coulombe PA (2012) Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. Nat Struct Mol Biol 19(7):707–715

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252(5009):1162–1164

MacPhee CE, Woolfson DN (2004) Engineered and designed peptide-based fibrous biomaterials. Curr Opinion Solid State Mater Sci 8(2):141–149

Meier M, Padilla GP, Herrmann H, Wedig T, Hergt M, Patel TR, Stetefeld J, Aebi U, Burkhard P (2009) Vimentin coil 1A - A molecular switch involved in the initiation of filament elongation. J Mol Biol 390(2):245–261

Mencarelli C, Ciolfi S, Caroti D, Lupetti P, Dallai R (2011) Isomin: a novel cytoplasmic intermediate filament protein from an arthropod species. BMC Biol 9:17

Mucke N, Wedig T, Burer A, Marekov LN, Steinert PM, Langowski J, Aebi U, Herrmann H (2004) Molecular and biophysical characterization of assembly-starter units of human vimentin. J Mol Biol 340(1):97–114

Nicolet S, Herrmann H, Aebi U, Strelkov SV (2010) Atomic structure of vimentin coil 2. J Struct Biol 170(2):369–376

Parry DAD (2006) Hendecad repeat in segment 2A and linker L2 of intermediate filament chains implies the possibility of a right-handed coiled-coil structure. J Struct Biol 155(2):370–374

Parry DAD, Fraser RD, Squire JM (2008) Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J Struct Biol 163(3):258–269

Patterson AL (1935) A direct method for the determination of the components of interatomic distances in crystals. Zeitschrift für Kristallographie-Crystalline Materials 90(1):517–542

Pekny M, Lane EB (2007) Intermediate filaments and stress. Exp Cell Res 313(10):2244–2254

Peter A, Stick R (2015) Evolutionary aspects in intermediate filament proteins. Curr Opin Cell Biol 32:48–55

Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol 9:51

Potapov V, Kaplan JB, Keating AE (2015) Data-driven prediction and design of bZIP coiled-coil interactions. PLoS Comput Biol 11(2):e1004046

Rämisch S, Lizatović R, André I (2015) Automated de novo phasing and model building of coiled-coil proteins. Acta Crystallogr D Biol Crystallogr 71(3):606–614

Robert A, Hookway C, Gelfand VI (2016) Intermediate filament dynamics: what we can see now and why it matters. BioEssays 38(3):232–243

Rossmann MG (1990) The molecular replacement method. Acta Crystallogr A: Found Crystallogr 46(2):73–82

Ruan J, Xu C, Bian C, Lam R, Wang JP, Kania J, Min JR, Zang JY (2012) Crystal structures of the coil 2B fragment and the globular tail domain of human lamin B1. FEBS Lett 586(4):314–318

Satelli A, Li S (2011) Vimentin in cancer and its potential as a molecular target for cancer therapy. Cell Mol Life Sci 68(18):3033–3046

Scally SW, Petersen J, Law SC, Dudek NL, Nel HJ, Loh KL, Wijeyewickrema LC, Eckle SBG, van Heemst J, Pike RN, McCluskey J, Toes RE, La Gruta NL, Purcell AW, Reid HH, Thomas R, Rossjohn J (2013) A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. J Exp Med 210(12):2569–2582

Snider NT, Omary MB (2014) Post-translational modifications of intermediate filament proteins: mechanisms and functions. Nat Rev Mol Cell Biol 15(3):163–177

Steinert PM, Roop DR (1988) Molecular and cellular biology of intermediate filaments. Annu Rev Biochem 57:593–625

Steinert PM, Marekov LN, Parry DAD (1993) Diversity of intermediate filament structure. Evidence that the alignment of coiled-coil molecules in vimentin is different from that in keratin intermediate filaments. J Biol Chem 268(33):24916–24925

Steinert PM, Marekov LN, Parry DAD (1999) Molecular parameters of type IV α-internexin and type IV-type III α-internexin-vimentin copolymer intermediate filaments. J Biol Chem 274(3):1657–1666

Strelkov SV, Burkhard P (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J Struct Biol 137(1–2):54–64

Strelkov SV, Herrmann H, Geisler N, Lustig A, Ivaninskii S, Zimbelmann R, Burkhard P, Aebi U (2001) Divide-and-conquer crystallographic approach towards an atomic structure of intermediate filaments. J Mol Biol 306(4):773–781

Strelkov SV, Herrmann H, Geisler N, Wedig T, Zimbelmann R, Aebi U, Burkhard P (2002) Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. EMBO J 21(6):1255–1266

Strelkov SV, Schumacher J, Burkhard P, Aebi U, Herrmann H (2004) Crystal structure of the human lamin A coil 2B dimer: implications for the head-to-tail association of nuclear lamins. J Mol Biol 343(4):1067–1080

Szeverenyi I, Cassidy AJ, Chung CW, Lee BT, Common JE, Ogg SC, Chen H, Sim SY, Goh WLR, Ng KW, Simpson JA, Chee LL, Eng GH, Li B, Lunny DP, Chuon D, Venkatesh A, Khoo KH, McLean WHI, Lim YP, Lane EB (2008) The Human Intermediate Filament Database: comprehensive information on a gene family involved in many human diseases. Hum Mutat 29(3):351–360

Taylor GL (2010) Introduction to phasing. Acta Crystallogr D Biol Crystallogr 66(Pt 4):325–338

Toivola DM, Strnad P, Habtezion A, Omary MB (2010) Intermediate filaments take the heat as stress proteins. Trends Cell Biol 20(2):79–91

Vincent TL, Green PJ, Woolfson DN (2013) LOGICOIL—multi-state prediction of coiled-coil oligomeric state. Bioinformatics 29(1):69–76

Wickert U, Mucke N, Wedig T, Muller SA, Aebi U, Herrmann H (2005) Characterization of the in vitro co-assembly process of the intermediate filament proteins vimentin and desmin: mixed polymers at all stages of assembly. Eur J Cell Biol 84(2–3):379–391

Wood CW, Bruning M, Ibarra AÁ, Bartlett GJ, Thomson AR, Sessions RB, Brady RL, Woolfson DN (2014) CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinformatics 30:3029–3035 btu502

Chapter 7
Lessons from Animal Models of Cytoplasmic Intermediate Filament Proteins

Jamal-Eddine Bouameur and Thomas M. Magin

Contents

Abstract Cytoplasmic intermediate filaments (IFs) represent a major cytoskeletal network contributing to cell shape, adhesion and migration as well as to tissue resilience and renewal in numerous bilaterians, including mammals. The observation that IFs are dispensable in cultured mammalian cells, but cause tissue-specific, life-threatening disorders, has pushed the need to investigate their function in vivo. In keeping with human disease, the deletion or mutation of murine IF genes resulted in highly specific pathologies. Epidermal keratins, together with desmin, are essential to protect corresponding tissues against mechanical force but also participate in

J.-E. Bouameur (✉) • T.M. Magin (✉)
Sächsischer Inkubator für Klinische Translation (SIKT) & Institute of Biology
Division of Cell and Developmental Biology, University of Leipzig,
Philipp-Rosenthal-Straße 55, 04103 Leipzig, Germany
e-mail: jamal-eddine.bouameur@uni-leipzig.de; thomas.magin@uni-leipzig.de

stabilizing cell adhesion and in inflammatory signalling. Surprisingly, other IF proteins contribute to tissue integrity to a much lesser extent than anticipated, pointing towards their role in stress situations. In support, the overexpression of small chaperones or the interference with inflammatory signalling in several settings has been shown to rescue severe tissue pathologies that resulted from the expression of mutant IF proteins. It stills remains an open issue whether the wide range of IF disorders share similar pathomechanisms. Moreover, we lack an understanding how IF proteins participate in signalling processes. Now, with a large number of mouse models in hand, the next challenge will be to develop organotypic cell culture models to dissect pathomechanisms at the molecular level, to employ Crispr/Cas-mediated genome engineering to optimize models and, finally, to combine available animal models with medicinal chemistry for the development of molecular therapies.

Keywords Intermediate filaments • Keratins • Epithelia • Vimentin • Desmin • GFAP • Peripherin • Syncoilin • Neurofilaments • α-internexin • Nestin • Synemin • Phakinin • Filensin • Animal models

List of Abbreviations

AMP   adenosine monophosphate
ATP   adenosine triphosphate
AxD   Alexander disease
BCC   basal cell carcinoma
CFTR   cystic fibrosis transmembrane conductance regulator
DAMPs   danger-associated molecular patterns
E   embryonic day
EBS   epidermolysis bullosa simplex
EGFP   enhanced green fluorescent protein
GFAP   glial fibrillary acidic protein
GLUT   glucose transporter
IF   intermediate filament
IFN   interferon
IL-   interleukin-
K   keratin
KI   knock in
KO   knock out
KtyI   keratin type I
KtyII   keratin type II
MDBs   Mallory-Denk bodies
MECD   Meesmann epithelial corneal dystrophy
MHC   major histocompatibility complex

NF-	neurofilament-
NF-κB	nuclear factor-kappa B
ROS	reactive oxygen species
Tg	transgenic
TNF	tumour necrosis factor
TNFR	tumour necrosis factor receptor
TPA	12-O-tetradecanoylphorbol-13-acetate
TSLP	thymic stromal lymphopoietin
WT	wild type

7.1 Introduction

Intermediate filaments (IFs) represent a super-family of fibrous protein assemblies in the animal kingdom (Eriksson et al. 2009). The family is divided into six sub-groups numbered I, II, III, IV, V and VI and is classified according to sequence homology. Lamins belong to homology group V and are present in the nucleus, where they comprise major components of the nucleoskeleton of all metazoans (Gruenbaum and Foisner 2015). Evolutionarily, lamins are considered as founding members of cytoplasmic IFs that originated most probably via gene duplication and deletion events (Peter and Stick 2015; Hering et al. 2016). All other IFs are localized in the cytoplasm; for space restrictions, we will focus here on data and concepts gathered from animal models of cytoplasmic IFs. Together with microtubules and microfilaments, IFs form major cytoplasmic networks that determine cytoarchitecture, provide resilience to mechanical force and regulate cell adhesion and cell migration (Goldman et al. 1996; Herrmann et al. 2007). Moreover, by acting as scaffolds, IF proteins participate in cell signalling and organelle positioning. Expression of IF results in cell type-, tissue-specific and differentiation-dependent combinations of proteins, which explains their widespread application as markers of differentiated states during development and pathogenesis (Moll et al. 2008; Szeverenyi et al. 2008). Looking at the individual IF types:

– Type I and II IF (keratins; KtyI and KtyII) constitute the most comprehensive class and are primarily expressed in epithelia (Fig. 7.1)
– Type III IFs consist of vimentin, desmin, glial fibrillary acidic protein (GFAP), peripherin and syncoilin (sync1 and sync2), and are predominantly found in mesenchymal, muscle, glial and neuronal tissues, respectively (Fig. 7.1)
– Type IV IF comprise α-internexin, neurofilaments (NF-H, NF-L and NF-M), nestin (full-length, nestin-short and Nes-SΔ107-254) and synemin isoforms (H or α, M or β and L), mainly expressed in neuronal and muscle cells and tissues (Fig. 7.1)
– Type VI IFs include two eye lens IFs that are highly divergent compared to others; phakinin (also named BFSP2 or CP49) and filensin (also named BFSP1 or CP115) (Fig. 7.1)
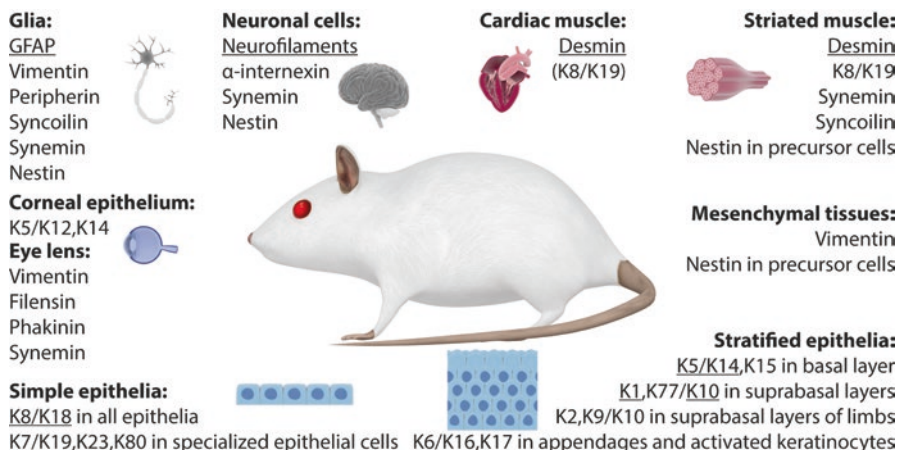
Fig. 7.1 Tissue repartition of intermediate filaments

Mutations in most IF genes can cause a wide range of tissue-restricted or systemic disorders, or lead to a predisposition to organ-specific diseases (Szeverenyi et al. 2008). Disruption of the cytoskeleton and occurrence of cytoplasmic or nuclear protein aggregates represent major features of IFopathies, in addition to altered growth and inflammatory signalling. In addition to elucidating basic and context-dependent functions of IF protein diversity, understanding the pathomechanisms of their corresponding disorders has been major driving force to establish mouse models with genetic deletions, alterations or transgenic expression of wild type (WT) and mutant IFs (Magin et al. 2004a, b). While many of these models faithfully reproduce aspects of human disease conditions, several of them have provided unexpected insights into biological processes depending on IF proteins that, in future, will require mechanistic analysis at the molecular level.

The overall domain organization of IF proteins is conserved and some domains have been crystallized, paving the way to an understanding of their overall structure (Herrmann et al. 2009, Block et al. 2015, Chernyatina et al. 2015; Bunick et al. 2016). IF proteins share a tripartite domain organization with a central α-helical rod domain of conserved length, flanked by highly variable head and tail domains. The rods contain heptad repeats that mediate strong and parallel coiled-coil homo- or hetero-dimerization of subunits. Dimerization is rapidly followed by subsequent stepwise assembly into antiparallel tetramers, unit-length filaments and into long non-polar filaments. The rod domains form the filament backbone while the divergent head and tail sequences regulate filament elongation and radial compaction, respectively (Herrmann and Aebi 2004). While mechanisms governing the *in vivo* assembly processes are little understood (Windoffer et al. 2011), it is evident that post-translational modifications, predominantly of head and tail domain sequences, play a major role in its regulation (Eriksson et al. 1992, 2004; Sihag et al. 2007; Kuga et al. 2013; Snider and Omary 2014; Loschke et al. 2015). Formation of an IF cytoskeleton and its spatiotemporal interaction with membrane adhesion sites, the

cell nucleus and other cytoskeletal elements, among others, involves accessory proteins. Among these, the plakin protein family members (also named "cytolinkers"), co/chaperones, 14-3-3 proteins and gigaxonin are major contributors. Depending on the cell types and expression profiles, plakin family members can attach IFs to microfilaments, microtubules, focal contacts, hemidesmosomes and desmosomes (Bouameur et al. 2014a). Gigaxonin is a member of the BTB–Kelch family, is active as a ubiquitin ligase and is preferentially expressed in the nervous system, loss of which causes neurodegeneration with sensory-motor deficits and muscle atrophy (Dequen et al. 2008; Cleveland et al. 2009; Ganay et al. 2011; Mahammad et al. 2013; Bomont 2016).

Here, we will first summarize the phenotypical alterations that result from the genetic analysis of IF proteins in their major compartments. We will then discuss common and specific principles, with an emphasis on molecular mechanisms, when possible. Finally, we will discuss open questions and point out future research directions.

7.2 Analysis of Keratins

Human and mouse genomes encode 28 KtyI and 26 KtyII genes, which are predominantly expressed as pair-specific combinations of type I and type II keratin proteins in epithelial cells and tissues. K8 and K18 (type II and type I, respectively) are present in all embryonic and simple epithelia and represent the only keratins in hepatocytes, acinar and islet cells of the pancreas and in proximal tubular epithelial cells of the kidney. In all other simple epithelia, this pair is complemented by varying combinations of K7, K19, K20, K23 and K80, indicative of distinct stages of differentiation and serving as excellent markers for the state of epithelial differentiation. K23 expression is elevated in a subset of colorectal tumours with microsatellites, along with the transcription factors FOXQ1 and CEBPA, correlating with elevated cell proliferation (Roth et al. 2012).

Keratin expression and organization in stratified epithelia is more complex, possibly reflecting the exposure of these tissues to considerable mechanical and diverse other stresses. The epidermis is composed of four main layers, which emerge from the continuous renewal of precursor cells and ultimately form the epidermal barrier (Potten and Morris 1988; Tumbar et al. 2004; Cotsarelis 2006; Simpson et al. 2011). Attached to the extracellular matrix (ECM), through hemidesmosomes and focal adhesions, proliferative keratinocytes represent the basal layer of the epidermis (Simpson et al. 2011; Tsuruta et al. 2011). This compartment expresses the keratin pair K5/K14, organized in loose bundles that extend from hemidesmosomes and desmosomes throughout the cytoplasm of highly prismatic cells. Upon terminal differentiation, these keratinocytes disconnect from the ECM and become terminally differentiated, accompanied by the replacement of K5/K14 with K1 and K10. During this process, a dramatic flattening of the keratinocytes and a reorganization of tightly bundled keratins, which run parallel to the cell surface, occurs (Fuchs and Green 1980; Reichelt et al. 2001; Simpson et al. 2011). Depending on regional

differences in the epidermis, additional keratins become expressed. At sites of high mechanical strain, the keratin pair K1/K10 is supplemented by the expression of additional keratins K2e and K9 at the palms of hands and soles (Candi et al. 2005; Moll et al. 2008). Being among the first proteins expressed during cornification, K1/K10 are considered as an essential protein scaffold that directs the sequential deposition and crosslinking of cornified envelope proteins (Candi et al. 2005). Epidermal injury triggers the rapid induction of K6, K16 and K17 at the wound edge, at the expense of K1/K10, to repair tissue function. This correlates with striking changes in keratin organization, morphological changes and functional properties of so-called "activated" keratinocytes (Reichelt and Magin 2002; Wong and Coulombe 2003; Kim et al. 2006). K6, K16 and K17 are also expressed in hair follicles and the nails (De Berker et al. 2000). Epidermal stem cells are located in protected niches, such as the bulge of the hair follicle and are characterized by the expression of K15 (Lyle et al. 1998; Watt 1998; Goldstein and Horsley 2012). The vibrissae bulge harbours two types of slow cycling stem cells, characterized by the expression of K5/K15/K17/K19 and K5/K17, respectively, which display loose keratin bundles in the former and tight bundles in the latter configuration (Larouche et al. 2008). Notably, both subpopulations lack K14 expression (Larouche et al. 2008). In view of recent data, which showed that differentiating keratinocytes are also capable of expressing K15, which is mediated through PKC (protein kinase C) signalling, the role of K15 as a stem cell marker has come into question (Bose et al. 2013). Neurosensory Merkel cells, which localize close to basal keratinocytes within the skin, express "simple epithelial type" keratins K8, K18, K19 and K20 (Moll et al. 1995; Sidhu et al. 2005). In the subsequent sections, we will discuss to which extent recent data from the analysis of transgenic animals support the concept of isotype-specific and context-dependent keratin functions, and whether this provides novel insights into the pathomechanisms of keratinopathies.

7.2.1   Simple Epithelial Keratins

7.2.1.1   Expression and Architectural Functions of Simple Epithelial Keratins

To address the contribution of keratins to the functionality of adult simple epithelia, K7, K8, K18 and K19 were deleted. Abolishing the presence of keratin filaments, either by genetic deletion of K8 together with K18 and K19, or by deleting the entire KtyII gene cluster, revealed that embryonic epithelia are fully dependent on an intact keratin cytoskeleton from the onset of placenta formation ~E9.5 (Hesse et al. 2000; Magin et al. 1998; Tamai et al. 2000; Kroger et al. 2011; Vijayaraj et al. 2010; Vijayaraj et al. 2009). In analogy to embryonic epithelia, adult simple epithelia remain intact in the absence of a keratin cytoskeleton, provided they are not exposed to stress (Omary et al. 2009; Alam et al. 2013; Asghar et al. 2015). Induction of stress, including partial hepatectomy or pharmacological stress is, however, not

tolerated in the absence of a keratin cytoskeleton and is followed by an increase in proliferation, inflammation or apoptosis (Loranger et al. 1997; Omary et al. 2009).

The type II keratin K7 is expressed in several epithelia (Smith et al. 2002; Moll et al. 2008) but has not been linked so far to any human genetic disease. Except in the bladder urothelium, where a slight increase in proliferation was noted, K7 KO mice did not show any major defect (Sandilands et al. 2013; Asghar et al. 2015). The absence of a more severe phenotype probably reflects the compensatory ability of K8, wherever it is co-expressed with K7.

To investigate the interdependence of keratins and desmosomes in the earliest stages of mouse development, autofluorescent GFP-K8 knockin (KI) mice have been generated (Schwarz et al. 2015). In this setting, *de novo* produced keratin filaments arise from dot-like keratin structures occurring in close proximity to desmosomes of pre-implantation embryos. The function of K8 during embryonic development is mouse strain-dependent: a KO in FVB/N mice led to 50 % embryonic lethality and major defects in some postnatal epithelia, whereas its depletion in a C57BL/6 background caused embryonic lethality at E12.5 with trophoblast giant cell layer failure with involvement of maternal TNF promoting cell death (Baribault et al. 1993; Jaquemar et al. 2003). The rescue of the K8 KO phenotype by tetraploid WT extra-embryonic tissues or by depletion of maternal TNF, revealed the requirement of a functional keratin cytoskeleton for trophoblast giant cells function, including apoptosis resistance. The partial penetrance of the K8 deletion in FVB/N mice may be due to keratin redundancy, or hint toward yet unidentified genetic modifiers.

Adult K8 KO FVB/N mice developed phenotypes in several organs covered by simple epithelia, including the liver, the gut, the pancreas and the thymus, where K7 is absent (Sandilands et al. 2013; Asghar et al. 2015). In line with the fact that sequence variants of K8 and K18 represent risk factors for the development and progression of liver diseases in humans (Szeverenyi et al. 2008; Toivola et al. 2015), K8 KO mice showed mild hepatitis (Loranger et al. 1997) and susceptibility to liver injury (Toivola et al. 1998). Of note, dysregulation of keratins, either by overexpression or deletion of a single keratin gene, can cause pathology, as exemplified by the formation of Mallory-Denk bodies (MDBs), which are cytoplasmic protein aggregates, in alcoholic liver disease, in non-alcoholic steatohepatitis and in transgenic animal models (Strnad et al. 2013). The major constituents of MDBs include K8, K18, ubiquitin, and the autophagy receptor p62. Despite intense research, it remains unclear whether MDBs fulfill a bystander, protective or injury-promoting function in several liver diseases, but they serve as histological and potential progression markers (Zatloukal et al. 2007; Strnad et al. 2013). Consistently, overexpression of human or mouse K8 in mice increased liver susceptibility to present MDBs and to injury (Nakamichi et al. 2005; Ku and Omary 2006; Kucukoglu et al. 2014). Moreover, increased basolateral levels of the bile canaliculus ecto-ATPase, an apical marker (Ameen et al. 2001), was observed in K8 KO liver, suggesting a partly disturbed polarity. Similarly, in the gut, several apical membrane proteins were not detected and γ-tubulin lost its sub-apical localization at distinct sites of the villus enterocytes of the small intestine from K8 KO animals (Toivola et al. 2004).

Electrolyte transporters were also mislocalized in the colon, leading to colitis, hyperplasia and diarrhea. The animals also developed hyperplasia and inflammation in the colon (Baribault et al. 1994). In the pancreas, K8 KO mice presented dysfunction of the exocrine pancreas, also due to mislocalized GLUT2 glucose transporter in islet β-cells (Alam et al. 2013). Some architectural abnormalities of the thymus have also been reported (Odaka et al. 2013). Functional redundancy between K18 and K19 is also pointed out by the fact that, contrary to the double KO, their single KOs developed less severe phenotypes than for K8 deletion. K18 deletion in mice leads to the development of mild hepatitis, with the noticeable formation of K8 aggregates, thus confirming the predominant role of K8/K18 in liver hepatocytes (Magin et al. 1998; Toivola et al. 2001). Old K18 KO mice spontaneously developed steatohepatitis and an increased frequency of hepatocellular carcinoma (HCC), suggesting that loss of K18, together with metabolic changes, presents a risk factor for the development of HCC defects (Bettermann et al. 2016).

Elevated expression of K8 and K18 in cells expressing the ΔF508 mutation of cystic fibrosis transmembrane conductance regulator (CFTR) is associated with increased colocalization of CFTR and K18 close to the endoplasmic reticulum. Further data substantiated the direct binding of K8, but not K18, to the ΔF508 mutant in nucleotide-binding domain-1. ΔF508-expressing mice that were topically treated with K8-siRNA showed restored nasal potential difference, equivalent to that of WT mice (Colas et al. 2012). Similar to CFTR patients, mice carrying the ΔF508 genetic mutation in the *Cftr* gene displayed reduced bone formation and decreased bone mass. Patients with cystic fibrosis display low bone mass and alterations in bone formation. Mice carrying the ΔF508 genetic mutation in the *Cftr* gene display reduced bone formation and decreased bone mass. Of note, murine and human osteoblasts express Cftr and K8 at low levels. Deletion of K8 in ΔF508-CFTR mice corrected overactive NF-κB signalling and decreased Wnt/β-catenin signalling induced by the ΔF508-CFTR mutation in osteoblasts. Thus, genetic or pharmacologic targeting of K8 leads to correction of osteoblast dysfunctions, altered bone formation and osteopenia in ΔF508-CFTR mice, providing a therapeutic strategy targeting the K8-ΔF508-CFTR interaction to correct the abnormal bone formation and bone loss in cystic fibrosis (Le Henaff et al. 2016). A different study came to another result, reporting that K18 directly interacts with CFTR, stabilizing it at the cell surface (Duan et al. 2012). In K18 KO mice, CFTR expression was drastically altered in the villi of the duodenum and in the columnar epithelia of the gallbladder. The K18-dependent surface localization and function of CFTR is in line with an analogous function in the correct localization of GLUT1 and GLUT3 glucose transporters (Vijayaraj et al. 2009). This correlated with lower forskolin-induced short circuit currents. At present, it is difficult to resolve the differences between these studies.

As for K19, all its functions apparently can be fulfilled by K18 in simple epithelia (Stone et al. 2007) (but not in muscle, see below). K19 KO mice developed a decreased hepatic ductal reaction after feeding with 3,5-diethoxycarbonyl-1,4-dihydrocollidin, indicating its contribution to cell proliferation in conjunction with Notch2 signalling (Chen et al. 2015). K20 overexpression using its own promoter caused no obvious phenotype, although overexpression of an artificial K20[R80H]

mutant, analogous to an epidermal keratin mutation causing epidermolysis bullosa simplex (see below), disorganized keratin filaments in the small intestine (Zhou et al. 2003).

Deletion of the entire KtyII gene cluster led to the absence of keratin filaments in all embryonic epithelia (Vijayaraj et al. 2009; Vijayaraj et al. 2010; Kroger et al. 2011). This was accompanied by the downregulation of protein biosynthesis through a mechanism that involves mislocalization of apical glucose transporters GLUT1 and GLUT3, and activation of AMP kinase, indicative of metabolic stress. The phosphorylation of the mTOR regulator Raptor by AMP kinase resulted in the subsequent down-regulation of protein biosynthesis in embryonic mouse tissues. Moreover, the amount of desmosomes was decreased in the absence of keratin IF. However, no major adhesion defects were notable in early embryonic epithelia (Vijayaraj et al. 2009). Further, yolk sac hematopoiesis and vasculogenesis, dependent on hedgehog/BMP4 signalling, were abnormal and reduced in the absence of keratin filaments, possibly because of decreased adhesion between endodermal and mesodermal cell layers (Vijayaraj et al. 2009, 2010; Kroger et al. 2011). Keratin filament-deficient trophoblast giant cells showed hypoxia inducible factor 1α, in line with a decreased interaction with maternal placental tissue (Kroger et al. 2011). A genetic rescue of the keratin filament deficiency, which was performed in several independent models, restored the above defects and furthermore revealed that differentiation of simple epithelia was possible in the presence of a single pair of keratins (Toivola et al. 2008; Kroger et al. 2013; Kumar et al. 2015). The fact that either K8/K18 or K8/K19 filaments (the latter containing the naturally tailless K19) could sustain development of simple epithelia until birth without obvious defects, suggests a considerable plasticity of keratins, at least in simple epithelia (Kumar et al. 2015).

7.2.1.2 Scaffolding and Regulatory Functions of Simple Epithelial Keratins

Sequence variants of K8 and K18, including $K8^{G62C}$ or $K8^{R341H}$, have been linked to chronic liver disease, raising questions about the underlying pathomechanisms (Ku and Omary 2006; Guldiken et al. 2015). Given that keratins, like other IF proteins, are subject to extensive posttranslational modifications, the question arises whether such sequence variants cooperate with stress-activated kinases. Based on the observation that $K8^{G62C}$ or $K8^{R341H}$ abolished the phosphorylation of $K8^{S73}$ but increased the activation of stress-activated protein kinases, the hypothesis was formulated that keratin networks may serve as a protective "phosphate sponge", redirecting kinases away from pro-apoptotic targets (Ku and Omary 2006; Snider and Omary 2014; Guldiken et al. 2015). In support, $K8^{S74A}$ or $K18^{S53A}$ overexpressing mice were shown to be more susceptible to liver injury and to apoptosis (Ku and Omary 2006). An alternative view is that such mutations alter the keratin dynamics that are required during tissue homeostasis and regeneration. Notably, the above K8 liver disease-associated variants trigger keratin hyper-sumoylation with consequent diminished solubility. In contrast, modest sumoylation of WT K8 promotes solubility. Hence, conformational changes induced by natural keratin mutations and

extensive tissue injury result in hypersumoylation of K8, K18 and K19. This sequesters keratins in an insoluble compartment, thereby limiting their cytoprotective function (Snider et al. 2011). Transcriptome profiling and immuno-precipitation experiments also identified nicotinamide N-methyl transferase as a K8 partner that was downregulated in the presence of K8 mutants. It also appears to contribute to the increased liver susceptibility of K8 mutants (Jaeschke et al. 2012; Guldiken et al. 2015). K18 phosphorylation also regulates keratin reorganization during mitotic progression via its interaction with the 14-3-3ζ protein (Ku et al. 2002).

K8 and K18 are also subject to single O-linked N-acetylglucosamine (O-GlcNAcase) Ser/Thr glycosylation (Chou et al. 1992; Ku and Omary 1995). To investigate its functional significance, mutations of three glycosylated serine residues into alanines in the head domain of K18 were introduced. This caused significantly increased susceptibility to O-GlcNAcase inhibitor-induced liver and pancreatic injury and apoptosis in a Fas/Akt1-dependent manner. In fact, K18 glycosylation stimulated the interaction of K8 with Akt1 and its activation by phosphorylation, and thereby promoted cell survival (Ku et al. 2010). In support of a keratin-dependent sequestration model, K8/K18 were found to modulate Raf-1 activity by recruiting it via 14-3-3 proteins (Ku et al. 2004). Whether the increased sensitivity of K8 KO mouse hepatocytes to Fas receptor-induced apoptosis is due to decreased activity of ERK, JNK MAP kinases and NF-κB through a similar mechanism, remains to be analyzed (Gilbert et al. 2004; Lee et al. 2013). Likewise, the decreased hepatic ductal reaction observed in K19 KO mice upon pharmacological stress was due to a role for K19 in cell proliferation, which was dependent on Notch2 signalling (Chen et al. 2015). A role of simple epithelial keratins in limiting cell proliferation is in line with the increased number of proliferating cells observed in heterozygous and homozygous K8 KO colon crypts (Asghar et al. 2015).

The aforementioned mislocalization of glucose transporters in KtyII KO embryos or K8 KO pancreatic cells indicated an unanticipated role of keratins in the regulation of cellular metabolism and growth. In KtyII KO embryos, glucose homeostasis was therefore disturbed, leading to the activation of AMP kinase, which acts as the major cellular energy sensor (Hardie 2007) and inhibits protein biosynthesis via the mTOR pathway in the presence of high levels of AMP. In K8 KO pancreas, the mislocalization of glucose transporters coincided with an anomalous ultrastructure of endocrine pancreas β-cells, a reduced pancreatic insulin content and an altered glucose regulation (Alam et al. 2013). Consequently, the animals presented a decreased damage sensitivity and diabetes induced by streptozotocin. In liver and hepatoma cells, K8/K18 loss was reported to lead to a more efficient glucose uptake and conversion by increased levels of hexokinases (Mathew et al. 2013). In the colon of K8 KO mice, the transcription of mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase 2 was down-regulated, together with its transcriptional activator, suggesting several roles for colonocyte keratins in energy metabolism (Helenius et al. 2015). An earlier study had revealed that K8 gene ablation caused changes in mitochondrial protein composition, distribution and in the size of mitochondria. The decreased ATP content and an increased cytochrome c release might contribute to an enhanced apoptosis susceptibility of K8 KO mice (Tao et al. 2009). Major mouse models for the analysis of simple epithelial keratins are listed in Table 7.1.

Table 7.1 Major animal models of type I and II IFs in simple epithelia

Proteins	IF type	Major tissue expression	Genetic models	Lethality	Major phenotypes	References
All type I keratins	I	Epithelia	KO mice	Birth	Intact simple epithelia due to presence of K8/K18, no IF in other epithelia, fragility of stratified epithelia, impaired epidermal barrier, hair follicle defects, unstable desmosomes, increased cell proliferation, altered mitochondrial composition and function	Kumar et al. (2015)
All type II keratins	II	Epithelia	KO mice	E9.5	Growth retardation downstream of mislocalized GLUT1/3, increased AMPK and decreased mTORC1, giant trophoblast defects, impaired vasculogenesis	Vijayaraj et al. (2009), Vijayaraj et al. (2010), and Kroger et al. (2011)
		Epithelia	KO + transgenic K8 rescue	Birth	Intact simple epithelia due to presence of K8/K19, no IF in other epithelia, fragility of stratified epithelia, impaired epidermal barrier, hair follicle defects, unstable desmosomes, increased cell proliferation, altered mitochondrial composition and function	Kroger et al. (2013), Bar et al. (2014), and Kumar et al. (2015)
		Epithelia	Partial KO + transgenic K8 rescue	P12	Localized acanthosis, hyperkeratotic epidermis, increased inflammation	Bar et al. (2014)

(continued)

Table 7.1 (continued)

Proteins	IF type	Major tissue expression	Genetic models	Lethality	Major phenotypes	References
K7	II	Simple epithelia	KO mice		Increased cell proliferation in bladder urothelium	Sandilands et al. (2013)
K8	II	Simple epithelia	KO C57/BL6 mice	E12.5	Trophoblast giant cell layer failure	Baribault et al. (1993) and Jaquemar et al. (2003)
		Pancreas	KO and heterozygous FVB/N mice		Abnormal β-cell ultrastructure and reduced pancreatic insulin content, altered glucose regulation, delayed diabetes induced by streptozotocin, decreased sensitivity to streptozotocin -induced damage	Toivola et al. (1998), Caulin et al. (2000), Zatloukal et al. (2000), and Alam et al. (2013)
		Colon	KO FVB/N mice	50 % at E9.5	Female sterility, colonic hyperplasia (hyperproliferation), ulcerative colitis, rectal prolapse, diarrhea (altered electrolyte transport)	Baribault et al. (1993), Baribault et al. (1994), Ameen et al. (2001), Jaquemar et al. (2003), Toivola et al. (2004), and Asghar et al. (2015)
		Liver	KO FVB/N mice		Mild hepatitis, altered morphology, predisposition to various injuries	(Loranger et al. (1997), Toivola et al. (1998), and Toivola et al. (2001)

K8 or K18	II/I	Liver/pancreas	Human or mouse K8 or K18 transgenic mice		Formation of Mallory-Denk bodies, susceptibility to high-fat-diet-induced liver damage and to progressive alterations in the exocrine pancreas	Nakamichi et al. (2005), Toivola et al. (2008), and Kucukoglu et al. (2014)
		Liver	Human or mouse K8 or K18 mutants transgenic mice		hK18^{S53A}, hK8^{G62C}, K8^{S74A} or K8^{R341H}; increased susceptibility to liver injury and apoptosis. hK18^{D238E,D397E} and hK18$^{S30/31/49A}$; amplified Fas-mediated liver pancreatic damage	Ku and Omary (2006), Guldiken et al. (2015), Ku et al. (2010), and Weerasinghe et al. (2014)
K18/K19 or K8/19	I	Simple epithelia	Double KO mice	E9.5	No filaments, fragile trophoblast giant cell layer (similar to K8 KO and KtyII KO)	Hesse et al. (2000) and Tamai et al. (2000)

(continued)

Table 7.1 (continued)

Proteins	IF type	Major tissue expression	Genetic models	Lethality	Major phenotypes	References
K18	I	Simple epithelia	KO mice		Mild hepatitis, K8 aggregates	(Magin et al. 1998; Toivola et al. 2001)
			hK18^{R90C} transgenic mice		Phosphorylation-dependent-mild hepatitis and an increased susceptibility to hepatic stresses with the formation of Mallory-Denk bodies	Ku et al. (2007) and Kwan et al. (2015)
			KO plus hK18^{R90C} transgenic mice	E9.5 in the absence of K18	Homozygotes: No IF but presence of keratin aggregates in all epithelia, giant trophoblast fragility	Hesse et al. (2007)
					Heterozygotes: Rescue by one K18 allele despite ubiquitous persistence of keratin aggregates	
K19	I	liverprogenitor cells, biliary epithelial cells and muscle	KO mice		Liver; reduced Notch 2 signalling and ductular reaction and increased susceptibly to cholestatic liver disease development	Stone et al. (2007), Lovering et al. (2011), and Chen et al. (2015)
					Striated muscle: mildly myopathic, weaker susceptibility to injury	
K20	I	Simple epithelia-intestine	hK20^{R80H} transgenic mice		Defective filament organization, no obvious pathology	Zhou et al. (2003)

7.2.2 Epidermal Keratins

7.2.2.1 Expression and Architectural Functions of Epidermal Keratins

Analysis of K5, K14 and K15 in the Basal Epidermis

Basal, proliferative keratinocytes of stratified epithelia express K5 and K14, together with lower amounts of K15 in a subset of cells (Bose et al. 2013). Mutations in K5 or K14 genes cause the skin fragility disorder epidermolysis bullosa simplex (EBS), which is characterized by the presence of cytoplasmic keratin aggregates, cytolysis and skin blistering of variable severity (Coulombe et al. 2009). The notion that similar phenotypes are observed in K5 and K14 KO mice, and in transgenic mice expressing mutant keratins, supports the hypothesis that the basal epidermis requires an intact keratin cytoskeleton to maintain cell integrity and adhesion (Lloyd et al. 1995; Peters et al. 2001; Arin and Roop 2004; Magin et al. 2004a; Arin et al. 2011). The observation that the loss of K14 (type I) was less severe compared to that of K5 in mice, due to the presence of K15 (type I), suggested that compensatory upregulation of another keratin represents a route toward therapy of EBS. In support, application of sulphoraphane, an inducer of the antioxidant transcription factor nuclear factor erythroid-derived 2 related factor 2-dependent Nrf2, known to induce K17, in addition to many additional target genes, improved the condition of K14 KO mice (Kerns et al. 2007). While this study seemed to suggest compensatory redundancy, expression of a chimeric molecule, consisting of simple epithelial K8 with the K5 coil 1A and coil 2B segments in the basal epidermis of K5 KO mice, using a K14 promoter, but not of K8, only provided a partial rescue. Surviving mice presented with alopecia and inflamed hyperkeratotic, scaly skin, misshapen nuclei and mitochondria with altered ultrastructure (Alvarado and Coulombe 2014). The reasons for the inability of K8 to compensate for K5 loss is possibly due to the lower amount of the former, which reduces the overall amount of keratins in the basal epidermis, or the decreased stability of complexes with K14 (compared to K5), in addition to the absence of Cys residues in K8. The latter might reduce the ability to reorganize keratin filaments *in situ* by Cys-mediated crosslinking, possibly affecting micromechanical tissue properties (Feng and Coulombe 2015).

The molecular mechanisms by which mutations in keratin genes cause EBS and other keratinopathies are not completely understood. In particular, it is unclear whether disease phenotypes result from a loss or a gain of function (Coulombe and Lee 2012). Expression of the most frequent EBS mutant, K14[R125C], together with WT K5 in Drosophila, resulted in adult flies with blistered wings, whereas WT keratins formed an extensive cytoskeleton without adverse effects on fly physiology. This implies a gain of toxic function caused by keratin aggregates. Thus, this fly model represents a powerful tool to study primary mechanisms of filament assembly in a living animal and of EBS (Bohnekamp et al. 2015).

Analysis of Suprabasal Epidermal Keratins

Expression of K1/K10 coincides with the onset of stratification during embryo development (Ouellet et al. 1990; Byrne et al. 1994), complemented by K2, K9 and K77 in subsets of suprabasal keratinocytes. During terminal differentiation, keratins, in particular K1, become covalently crosslinked to the cornified envelope (Candi et al. 2005; Kumar et al. 2015). In analogy to K5 and K14, mutations in K1 and K10 genes give rise to localized or generalized epidermolytic or non-epidermolytic ichthyoses, frequently accompanied by loss of barrier integrity (Arin et al. 2011). Genetic ablation of the K1 gene in mice caused a major cornified envelope defect, accompanied by a slightly disturbed inside-out barrier. Otherwise, the epidermis of mice appeared largely unaffected, possibly due to upregulation of keratins K6 and K16. Despite the mild histological phenotype, loss of K1 caused early postnatal mortality with full penetrance, pointing to non-structural functions of K1 (see below) (Roth et al. 2012). At the same time, deletion of K10, the heterodimer partner of K1, did not compromise animal survival, indicating that even heterodimeric partner keratins perform non-overlapping functions. Mice doubly deficient for K1/K10 died shortly after birth and showed a fragile epidermis with smaller desmosomes and decreased amounts of desmosomal proteins, but were reported to have no significant barrier defects (Wallace et al. 2012). Thus, the low amounts of K5/K14 persistent in suprabasal layers are not enough to stabilize desmosomes *in vivo*, although they do so in cultured keratinocytes (Loschke et al. 2016).

In addition to K1, K76 is expressed in suprabasal epithelia of the epidermis, vagina, the eyelid, palate and gingiva (Moll et al. 2008; DiTommaso et al. 2014). KO of K76 resulted in non-blistered flaky skin, which decreased with the development of hair follicles, in addition to unkempt and dull coats and smaller body size (DiTommaso et al. 2014). KO animals survived up to 3 months and developed infections of skin wounds, which failed to properly resolve over time. Possibly, mislocalization of the tight junction protein claudin-1 might have contributed to the noted epidermal barrier defects. In another K76 KO mouse model, hyperplastic lesions in oral epithelia were reported (Ambatipudi et al. 2013).

In mouse foot soles, K2 and K10 are the most prominent suprabasal keratins with the exception of the footpads where K1, K9 and K10 predominate (Fischer et al. 2014, 2015). In fact, the two KtyII – K1 and K2 – are differentially expressed in the suprabasal epidermis. In back skin and similar to humans, K1 is the major type II keratin in the interfollicular epidermis while K2 is detected in ear, interscale regions of tail skin and some regions of the sole and toe pads, except in eccrine sweat glands. Therefore, in ears and the tails, K1 is replaced by K2 in the interfollicular area, while K1 is restricted to infundibular regions. This alternating pattern is also found in the plantar epidermis. Formation of cytoplasmic aggregates upon deletion of K1 or K2 strongly indicates that K10 is their primary heterodimer partner (Fischer et al. 2014, 2015). Unlike the lethal phenotype of K1, K2 KO mice were mildly affected and developed more pigmented, scaly skin on the ears and hyperkeratotic calluses on the soles and toe pads with hyperkeratosis. In addition, mild barrier

defects accompanied by localized inflammation were observed in the ears of K2 KO animals, with increased production of TSLP and IL-18 and infiltrates of mast and T cells. Mice lacking both K2 and K10 were viable and developed epidermal acanthosis and hyperkeratosis in inter-footpad epidermis of the soles. Their survival may be due to the more restricted expression of K2 compared to K1 and the highly elevated expression of K1, K9 and K16 in the soles (Fischer et al. 2015).

In humans, K9 mutations cause epidermolytic palmoplantar keratoderma. KO of K9 resulted in viable mice, which presented calluses with hyperpigmentation of stress-bearing footpads (Fu et al. 2014). Without K9, K5 and K2 were significantly decreased, while the stress keratins K6 and K16 were strongly induced (Fu et al. 2014).

Analysis of Keratins Expressed in Corneal Epithelia

K3/K12 forms the main keratin filaments of the human corneal epithelium and their mutations in human are associated with corneal dystrophies with microcyst formation and cytolysis including Meesmann epithelial corneal dystrophy (MECD) (Irvine et al. 1997). As K3 is absent in mice, K5 has been shown to be the major partner of K12 in the corneal epithelium (Lu et al. 2006). In the absence of K5, however, K4 can compensate the loss. Given that K12 KO mice suffered from mild corneal epithelial erosion (Kao et al. 1996), a K12^{L132P} KI mouse model was generated to study the pathomechanisms relevant for MECD (Allen et al. 2016). In homozygous mice, cytolysis and corneal fragility, accompanied by detachment of the epithelium at the stromal interface was noted, heterozygotes displayed a milder phenotype. At the mRNA level, K14, K4 and K10 were up- and K5 was down-regulated. Notably, the mutant K12 formed cytoplasmic protein aggregates, which triggered the unfolded protein response in order to prevent cell death (Nakagawa et al. 2000; Allen et al. 2016). Activation of the unfolded protein response was able to prevent apoptosis in heterozygous K12-mutant mice, but failed to do so in the homozygotes (Allen et al. 2016). The molecular mechanisms inducing an unfolded protein response and their potential contribution to epidermal disorders, also accompanied by aggregated keratins, require future investigation.

7.2.2.2 Regulatory Functions of Epidermal Keratins

In addition to major structural defects caused by mutations of epidermal keratins, evidence is accumulating to support the involvement of a subset of keratins in cell growth and inflammation.

Absence of K10 in mice caused no major structural skin defects, except that ear and tail skin displayed hyperkeratosis and acanthosis with increased keratinocyte proliferation, accompanied by the suprabasal induction of c-Myc, cyclin D1, 14-3-3Σ, p38 and ERK1/2 phosphorylation (Reichelt and Magin 2002; Reichelt et al. 2004). To test whether K10 suppresses cell proliferation and tumour formation in the skin, as was previously suggested (Paramio et al. 2001; Santos et al. 2002), a

two-step skin carcinogenesis model was employed. K10 KOs developed far less papillomas than WT mice (Reichelt et al. 2004). Analysis of label-retaining cells revealed a strongly accelerated keratinocyte turnover in K10 KO epidermis, suggesting an increased elimination of initiated keratinocytes at early stages of developing tumours. Despite a strong upregulation of K6, K16 and K17 in K10 null epidermis, no obvious inflammation was observed. In a second experiment, a chimeric keratin that consisted of the K14 rod domain fused to the K10 head and tail domains (K1014chim) was expressed in the basal epidermis (Chen et al. 2006). The K10 end domains had no effect on basal keratinocyte proliferation *in vivo*. Following the aforementioned chemical skin carcinogenesis protocol, papilloma formation in mutant mice was accelerated instead of being inhibited. In this setting, increased tumour susceptibility of K1014chim mice was considered to partly result from a suppression of apoptosis. In many carcinomas, tumour cells invade as a cohesive multicellular unit, a process termed collective invasion. In primary breast tumours, collective invasion was led by a subset of tumour cells that sustained high levels of K14. Abolishing the expression of K14 disrupted collective invasion (Cheung et al. 2013). Thus, in contrast to predictions from cell culture models, expression of keratins can contribute to tumour invasion, possibly by strengthening cell adhesion and providing mechanical integrity (Kroger et al. 2013; Seltmann et al. 2013b).

As mentioned before, K1 KO mice died perinatally, despite minor skin defects (Roth et al. 2012). Transcriptome profiling of prenatal mouse epidermis revealed a strong upregulation of pro-inflammatory cytokines and of S100A8/9 proteins, representing members of the alarmin family of danger-associated molecular patterns (DAMPs). Overall, K1 KO mouse epidermis shared a gene expression signature with atopic eczema or psoriatic epidermis, with the notable elevation of pro-inflammatory cytokines including IL-18 or TSLP but not IL-1α and TNFα. Pharmacological inhibition of IL-18 or the generation of double KO animals for K1 and IL-18, partially rescued the K1 phenotype. The involvement of K1 in the release of IL-18 was further corroborated by K1 knockdown in cultured keratinocytes. This resulted in increased IL-18 processing and release after caspase-1 cleavage, regulated by the inflammasome (Davis et al. 2011; Roth et al. 2012). These data suggested that K1 participates in a keratinocyte-based network that restrains inflammation upon intracellular damage or barrier disruption.

In addition to the epidermis, K6a is detected in developing mammary glands, in a subset of mammary luminal epithelial and body cells of terminal end buds, but K6 ablation in mice did not impair their development (Grimm et al. 2006). Combined deletion of the K6a and K6b genes did not interfere with mammary gland development, but it was accompanied by an increase in the number of steroid-receptor-positive cells. In the epidermis, K6 is the preferred partner of K16 and K17 in appendages, in the innermost layers of the outer root sheath of hair follicles, and is induced in response to oxidative and other forms of stress or wounding (Wojcik et al. 2000; Kerns et al. 2007, 2010). K6a KO mice present a delayed skin wound healing, in parts through controlling the activity of Src kinase (Wojcik et al. 2000). Mutations of K6a, K6b and type I K16 and K17 cause pachyonychia congenita, a rare autosomal dominant disease with hypertrophic nail dystrophy, palmoplantar keratoderma and oral leukokeratosis (McLean et al. 2011).

K16 KO mice develop oral lesions and hyperkeratotic calluses in front and hind paws, reminiscent of palmoplantar keratoderma and of pachyonychia congenita (Lessard and Coulombe 2012). A gene expression analysis of affected tissues revealed a signature enriched in genes involved in inflammation and innate immunity, in particular damage-associated molecular patterns (DAMPs). Many of these genes are also upregulated upon barrier damage. A comparison of K16 KO and WT keratinocytes provided evidence for cell-intrinsic and K16-dependent upregulation of DAMPs upon phorbol-ester-mediated stimulation. The remaining challenge will be to identify the mechanism by which K16 regulates a pathway that is activated when the skin experiences barrier defects and increased or altered mechanical forces, e.g., upon wounding. Remarkably, similar pachyonchyia-like alterations were found in K16-overexpressing mice (Presland et al. 2004). In these, tyrosine phosphorylation of the EGF receptor was increased in newborn skin (Paladini and Coulombe 1998). In an attempt to shed light onto the mechanisms by which loss of K16 contributes to hyperkeratosis and inflammation, a recent study identified K16 as a negative regulator of Nrf2, limiting activation of major Nrf2 target genes. Conversely, topical application of the NRF2 activator sulphoraphane to Krt16-/- mice prevented the development of PPK (Kerns et al. 2016).

K17 is constitutively expressed in hair follicles and epidermal appendages (Bianchi et al. 2005). In line with K17 expression in follicles, the development of a full pelage hair coat was slightly delayed and most of the vibrissae were missing in K17 KO mice (McGowan et al. 2002; Larouche et al. 2008), showing K17 involvement in hair follicle development and/or cycling rate. Early alopecia of K17 KO was associated with hair shaft fragility and apoptosis in the hair bulb cells due to defects of hair follicle cycling (Tong and Coulombe 2006). The mechanism by which K17 regulates hair cycling likely involves attenuation of a TNFα-dependent apoptotic response and probably NF-κB activity. In fact, the K17 KO phenotype was rescued by a double KO for K17 and TNFα. Mechanistically, K17 interacts directly with TNF receptor 1-associated death domain protein, a death adaptor during TNFR1-dependent signalling.

Moreover, K17 expression is highly induced in the epidermis during wound healing, upon barrier damage and in hyperproliferative and inflammatory conditions including psoriasis in an AP1- and MAPK-dependent manner (Kerns et al. 2007, 2010). The decreased size of K17 KO keratinocytes suggested its involvement in growth control. In due course, it was found that K17 promotes protein biosynthesis through interacting with 14-3-3Σ which represses mTORC1 activity by sequestration of TSC1/2 (Kim et al. 2006). Thus, K17 expression may coordinate cytoskeletal reorganization and growth requirements during wound healing. While constitutive expression of K17 in its natural compartments does not drive tumourigenesis, its overexpression in Gli2 transgenic mice that serve as a model for basal cell carcinoma (BCC), induced select Th1 cytokines with established roles in BCC and promoted tumourigenesis, accompanied by inflammation. Conversely, ablation of K17 in the same model delayed tumourigenesis, reduced inflammation and shifted the inflammatory cytokine expression from a Th1- and Th17-dominated to a Th2-dominated profile (Depianto et al. 2010). At the same time, the formation of HPV16-driven cervical squamous cell carcinoma lesions was delayed in K17 KO

mice and was accompanied by increased Th1-associated pro-inflammatory cyto-kines, immune cells and differentiation markers (Hobbs et al. 2016). First insights into underlying mechanisms come from two studies, stimulated by the observation that K17 is induced in basaloid skin tumours, coinciding with a poor prognosis (Wang et al. 2013). The investigation of K17-interacting proteins revealed that it can interact with the RNA-binding protein hnRNP K, resulting in its cytoplasmic local-ization. As a consequence, multiple pro-inflammatory mRNAs including those encoding the CXCR3 ligands CXCL9, CXCL10, and CXCL11, were upregulated, promoting cell growth and invasion. This supports a keratinocyte-intrinsic role of K17 in integrating the heterogeneous nuclear ribonucleoprotein hnRNP K-mediated control of CXCR3 signalling that promotes tumourigenesis (Chung et al. 2015). A second study pointed to an interaction of the transcription regulator Aire (Kumar et al. 2011) and K17 that may take place in the cell nucleus. In this compartment, K17 may stimulate transcription of Aire by interacting with hnRNP K. Further, K17 may act as a scaffold for the Aire protein to further promote Aire-dependent expres-sion of pro-inflammatory genes, possibly from NF-κB consensus sequences (Hobbs et al. 2015). It will be interesting to learn how this exciting scaffolding function of K17 in cytoplasmic and nuclear compartments is regulated to promote inflamma-tion only in a limited tissue context.

Deletion of the Entire Keratin Protein Family

The animal models discussed so far have shown that the integrity and stress responses of major compartments of stratified epithelia relies on the presence of an intact keratin cytoskeleton and that deletion or overexpression of a subset of kera-tins can additionally result in inflammatory and/or hyperproliferative conditions. In view of keratin redundancy, one needs to ask whether such conditions result from a loss of isotype-specific functions, from a misbalance of keratins upon deletion of a single gene or from tissue context-dependent mechanisms. To begin to address that, mice that lack the entire type I or type II keratin gene complement were generated, resulting in the absence of all keratin filaments in the epidermis (Kroger et al. 2013; Bar et al. 2014; Kumar et al. 2015).

   Both strains of mouse developed to term, showing that epidermal morphogenesis and overall differentiation do not depend on a keratin cytoskeleton. Both KtyI and KtyII KO mice displayed normal epidermal morphology at the onset of stratifica-tion (E15.5) but developed numerous abnormalities until E18.5 with extreme skin fragility that caused perinatal mortality with full penetrance. Transcriptome profil-ing of prenatal skin revealed the upregulation of distinct alarmins, the induction of the antioxidant transcription factor Nrf-2 and the dysregulation or epidermal barrier genes. In strong support of data from isolated, keratin-deficient keratinocytes, des-mosomal proteins were reduced in their amounts and accumulated in the cytoplasm, demonstrating a role of keratins in desmosome maintenance. Additional data indicated that keratins control desmosomes in part by sequestering PKC-α through interaction of the Rack1 scaffold (Kroger et al. 2013; Loschke et al. 2016). Beyond a major role in support of desmosome adhesion, two major phenotypes

were uncovered in keratin-deficient mice. Following the observation that cornified envelopes were disrupted in the absence of keratins, their composition was analyzed by proteomics. This showed that early cornified envelope assembly steps were less dependent on keratins whereas incorporation and crosslinking of filaggrin, loricrin and hornerin were impaired in the absence of keratins. Further data showed that in this setting, Nrf-2-dependent mechanisms failed to restore a functional barrier, suggesting that keratins, in particular K1 and K10, not only serve as cornified envelope constituents but also as structural scaffolds to mediate stepwise incorporation of major components (Kumar et al. 2015). Consistent with barrier defects, increased inflammation prevailed in mice lacking K1, K16 or K17. Of note, mutations in major cornified envelope constituents including filaggrin, hornerin, desmoplakin and desmoglein 1 have been linked to atopic dermatitis or multiple allergies and metabolic wasting, all displaying barrier defects in humans (Palmer et al. 2006; Smith et al. 2006; Nestle et al. 2009; Oji et al. 2010; Samuelov et al. 2013; Matsui and Amagai 2015; McAleer et al. 2015). Thus, one can hypothesize that mutations in a subset of epidermal keratin genes may contribute to inflammatory skin disorders. Mechanistically, this may arise through barrier disruption due to the scaffolding function of keratins, in addition to cell-intrinsic events, as shown for K1, K16 and K17 (Roth et al. 2012; Lessard et al. 2013; Chung et al. 2015). A second and unexpected role of keratins resulted from the altered distribution of mitochondria in keratin-deficient mice. Subsequent studies showed a keratin-dependent and cell-intrinsic regulation of mitochondrial lipid and protein composition that was accompanied by increased oxygen consumption in the absence of keratins (Kumar et al. 2015). Whether keratins are directly involved in mitochondrial fusion and fission and in the interaction of mitochondria with the endoplasmic reticulum where most mitochondrial lipids are synthesized, remains to be analyzed. Collectively, these findings illustrate not only that pathomechanisms underlying keratinopathies are much more complex than previously suggested but also offer new avenues for improving disease conditions. Major mouse models for the analysis of epidermal keratins are listed in Table 7.2.

7.3 Analysis of Type III Intermediate Filament Proteins

7.3.1 Vimentin

Vimentin is widely expressed in cell and tissues of mesenchymal origin. On the basis of *in vitro* and cell culture studies, it was expected to fulfil vital functions in cell signalling, mechanics, adhesion and migration. Surprisingly, vimentin KO mice showed unimpaired development and lifespan (Colucci-Guyon et al. 1994; Eckes et al. 2000; Ivaska et al. 2007). Of note, mesoderm formation is unaffected by the absence of vimentin. Under stress conditions, for instance upon skin wounding in embryonic or adult epidermis, vimentin KO mice showed a delay in wound closure (Eckes et al. 2000), connected to an altered differentiation of dermal fibroblasts into myofibroblasts, which impacts on cell migration. In line, vimentin KO blood

Table 7.2 Major animal models of type I and II IFs in stratified epithelia

Proteins	IF type	Major tissue expression	Genetic models	Lethality	Major phenotypes	References
K1	II	Suprabasal layer	KO mice	Shortly after birth	Barrier defect, minor tissue fragility, induction of innate immune response, upregulation of IL-18, transcriptome profile related to atopic dermatitis	Roth et al. (2012)
			K1/IL-18 double KO mice	Normal life span	Partial rescue by genetic deletion of IL-18	Roth et al. (2012)
K1/K10	II/I	Suprabasal layer	K1/K10 double KO mice	Shortly after birth	Severe skin fragility, desmosomal defects	Wallace et al. (2012)
K2	II	Suprabasal layer (sole, toe pad and ears)	KO mice		More pigmented scaly skin on the ears and hyperkeratotic calluses on the soles and toe pads with hyperkeratosis. Mild epidermal barrier defects accompanied by localized inflammation in the ears	Fischer et al. (2014)
			K2/K10 double KO		Almost similar to K2 KO	Fischer et al. (2015)
K5	II	Basal layer	KO mice	Shortly after birth	Extremely fragile skin with cytolysis at basal layer	Peters et al. (2001)

K5/K8	II		K5/K8 chimera KI mice	Partial rescue of K5 KO phenotype, alopecia, locally inflamed and hyperkeratotic, scaly skin, aberrant nuclei and mitochondria	Alvarado and Coulombe (2014)
K5/K14	II/I	Most fly tissues	Ectopic expression in Drosophila Transgenic hK5/hK14	Cytoplasmic keratin network, fully compatible with fly development and physiology	Bohnekamp et al. (2015)
			hK5/K14^{R125C} ectopic expression in transgenic drosophila	Wing blisters, widespread keratin aggregates, adhesion defects between epidermis and fly wing heart epithelium	Bohnekamp et al. (2015)
K6	II	Hair follicles and activated keratinocytes	KO mice	Delayed skin wound healing	Wojcik et al. (2000) and Grimm et al. (2006)
K9	I	Suprabasal layer of palmoplantar epidermis	KO mice	Calluses with hyperpigmentation of stress-bearing footpads, keratotic phenotypes with hyperproliferation and disrupted cytoskeletal integrity in localized area of the footpads	Fu et al. (2014)

(continued)

Table 7.2 (continued)

Proteins	IF type	Major tissue expression	Genetic models	Lethality	Major phenotypes	References
K10	I	Suprabasal layers	KO mice		Normal epidermis, hyperkeratosis in ear and tail skin, reduced tumour numbers and size in two-stage carcinogenesis model	Reichelt et al. (2001), Reichelt and Magin (2002), and Reichelt et al. (2004)
			Partial KO > K10 T mutant	Shortly after birth	Epidermolytic hyperkeratosis	Reichelt et al. (1997), Reichelt et al. (1999)
			K10^{R154C} KI mice		Epidermolytic hyperkeratosis-like blister; scaling, hyperkeratosis, stratum spinosum desintegration	Arin and Roop (2004)
K12	I	Corneal epithelium	KO mice		Mild corneal epithelial erosion	Kao et al. (1996)
			Human K12^{L132P} KI mice		Fragile corneal epithelium, induction of unfolded protein response and apoptosis	Allen et al. (2016)
K14	I	Basal layer	KO mice	Normal if survive the first weeks	Epidermal fragility with cytolysis at the basal layer, improvement under K17 induction by sulphorophane treatment	Lloyd et al. (1995) and Kerns et al. (2007)
			K14^{R131C} KI mice	Shortly after birth	Epidermolysis bullosa simplex-like phenotype	Arin and Roop (2004)
			K10/K14 chimera mice		Increased papilloma formation when subjected to chemical skin carcinogenesis	Chen et al. (2006)

K16	I	Appendage epidermis, hair follicles	KO mice	Oral leukokeratosis, hypertrophic nail dystrophy, palmoplantar keratoderma with increased alarmin expression, decreased oxidative stress and NRF2 response	Lessard and Coulombe (2012), Lessard et al. (2013) and Kerns et al. 2016
			hK16 transgenic mice	Similar to KO, hyperproliferation, inflammation	Presland et al. (2004)
K17	I	Appendage epidermis, hair follicles	K17 KO mice	Delayed development of a full pelage hair coat, absence of vibrissae; attenuated hyperplasia and inflammation in TPA-induced acute dermatitis-like phenotype; attenuated skin tumourigenesis in HPV16Tg-induced and Gli2-dependent	McGowan et al. (2002), Larouche et al. (2008), Depianto et al. (2010), and Hobbs et al. (2015; 2016)
K75 (K6hf)	II	Stratified epithelia	N159del mutant mice	Pachyonychia congenita-like	Chen et al. (2008)
K76	II	Stratified epithelia-epidermis, vagina, the eyelid, palate and gingiva	KO mice	Hyperplastic lesions in oral epithelium	Ambatipudi et al. (2013)
			KO mice	3 months	Non-blistered flaky skin which decreases with the development of hair follicles, defective tight junction	DiTommaso et al. (2014)

mononuclear cells showed a markedly reduced capacity to home to mesenteric lymph nodes and spleen in mice. Furthermore, direct interaction of vimentin with integrin β3 underneath the plasma membrane is thought to positively contribute to migration and metastasis of a murine breast cancer cell line injected in BALB/c mice by regulating integrin-ligand interaction (Kim et al. 2016). Also, endothelial integrity was compromised in vimentin KO mice, demonstrating a vimentin contribution to the barrier that governs leukocyte extravasation. A recent study reported significantly delayed skin wound healing in Vim-/- mice, characterized by reduced fibroblast growth, accompanied by diminished activation of TGF-ß1 and slug, to major EMT activators. Thus, vimentin appears to coordinate wound healing by controlling fibroblast proliferation, TGF-β1– and Slug signalling (Cheng et al. 2016). Absence of vimentin resulted in highly aberrant expression and distribution of surface molecules critical for homing. These data show that IFs are active in lymphocyte adhesion and transmigration. Vimentin KO mice challenged with intraperitoneal *Escherichia coli* control the bacterial infection better than WT. *In vitro*, vimentin KO phagocytes showed significantly increased capacity to mediate bacterial killing by abundant production of reactive oxygen species and nitric oxides, likely due to interactions with the p47phox active subunit of NADPH oxidase. In a model of acute colitis induced by dextran sodium sulphate, vimentin KO mice developed significantly less gut inflammation than do WT mice. These data suggested that vimentin impedes bacterial killing and production of reactive oxygen species and thereby may contribute to the pathogenesis of acute colitis (Mor-Vaknin et al. 2013). In the context of lung injury, activation of the NLRP3 inflammasome and maturation of IL-1β result in inflammation and fibrosis. Upon stimulation of lungs with LPS, bleomycin and asbestos, vimentin KO mice showed decreased caspase-1 and IL-1β levels. The direct protein-protein interaction between NLRP3 and vimentin suggests that vimentin may be a regulator of the NLRP3 inflammasome in a manner opposite to that of K1 (Roth et al. 2012; dos Santos et al. 2015).

To investigate whether vimentin might be involved in disease, a transgenic mouse model expressing the dominant mutation Arg113Cys (analogous to K14 Arg125Cys in EBS) in vimentin was generated, using the endogenous vimentin promoter. Despite the presence of cytoplasmic vimentin aggregates in all vimentin-expressing tissues, pathology was restricted to the eye lens where posterior cataracts in lens fibre cells formed (Bornheim et al. 2008). The prediction that mutations in the vimentin gene might cause cataracts, was confirmed soon after (Muller et al. 2009).

A variety of protein kinases have been identified to phosphorylate vimentin in a spatiotemporal regulated manner (Sihag et al. 2007). Among the resulting phenotypes, preventing vimentin phosphorylation during mitosis exerts the most dramatic phenotype, inhibiting cytokinesis by the retention of an IF bridge-like structure connecting the two daughter cells (Yatsunami et al. 1991; Yasui et al. 1998, 2001; Goto et al. 2003; Kawajiri et al. 2003; Yamaguchi et al. 2005). In a tour-de-force, vimentin KI (Vim[SA] KI) mice, in which 11 Ser residues in the vimentin head domain known to be phosphorylated were mutated to Ala, were generated (Matsuyama et al. 2013). Vim[SA] KI mice developed microophthalmia 2 months after birth, followed by a cataract at the age of 1 year. Importantly, lens epithelial cells of Vim[SA/SA] KI mice accumulated bi-nucleated, abnormal numbers of centrosomes per cell and aneuploid

nuclei, indicative of chromosomal instability. The failure of Vim[SA] KI mice to undergo cytokinesis resulted in tetraploidy, accompanied by elevated mRNA levels of the cell cycle inhibitor p21 and an increase in aneuploidy. The age-dependent loss of subcutaneous fat in Vim[SA] KI mice suggested that a subset of tetraploid cells entered a new cell cycle and developed into aneuploid cells *in vivo*, which promoted premature ageing (Tanaka et al. 2015). Collectively, these data are consistent with a major requirement of vimentin phosphorylation during cell cycle regulation and cytokinesis in vimentin-expressing cells.

7.3.2 *Desmin*

Desmin represents the major constituent of IFs in striated and cardiac muscle which localize around Z-discs and extend to organelles. It connects Z-lines to costameres at the sarcolemma via cytolinkers like plakins and dystrophin and to intercalated discs in cardiac muscles (Lieber et al. 1994; Li et al. 1997; Capetanaki et al. 2007; Winter and Wiche 2013). Mutations in the single desmin gene can cause myofibrillar myopathies and dilated cardiomyopathies. Desmin KO mice revealed major defects in striated and cardiac muscle, supporting its vital role (Li et al. 1996; Milner et al. 1996, 1999; Mavroidis and Capetanaki 2002, Hnia et al. 2015). KO mice developed skeletal myopathy with increased tetanic force production, decreased muscle stiffness and coupling between longitudinal and transverse properties and dilated cardiomyopathy followed by extensive loss of cardiomyocytes. Accompanying fibrosis and calcification impaired systolic function and ultimately led to heart failure (Li et al. 1996; Milner et al. 1996, 1999; Boriek et al. 2001; Fountoulakis et al. 2005). At the cellular level, desmin KO myotubes lose costameric structures in fast-twitch muscles (with loss of contractile material, loss of proper myofibrillar alignment, and separation of myofibrils from the intercalated disc) (O'Neill et al. 2002), correlating with less generated force and an increased susceptibility to damage after eccentric exercise (Li et al. 1997). Transgenic cardiac re-expression of desmin in desmin KO mice clearly demonstrated the importance of desmin in the heart phenotype and excluded possible involvement of coronary circulation that might cause myocardial abnormalities (Weisleder et al. 2004a). Fibrosis and dystrophic calcification in desmin-null myocardium could represent secondary defects caused by inflammation in the context of cardiomyocyte necrosis. In support, synthesis and deposition of extracellular matrix proteins osteopontin and decorin, activated complement elements, transforming growth factor-beta1 and of angiotensin-converting enzyme were increased (Mavroidis and Capetanaki 2002; Mavroidis et al. 2015). Remarkably, mice doubly deficient for desmin and the complement 5 receptor (C5aR) displayed a significantly improved cardiac phenotype. Improvement of the phenotype of desmin KO mice was also found by treatment with a pharmacological inhibitor of C5a signalling (Mavroidis et al. 2015). Additional defects in desmin-deficient mice were altered, including positioning of cell nuclei along muscle fibres and the number and morphology of mitochondria (Ralston et al. 2006; Goodall et al. 2012). Independent studies strongly suggested

that structural and functional alterations of mitochondria are a primary aftermath of plectin deficiency in muscle, contributing to myofibre degeneration. It appears that the plectin isoform P1b is involved in controlling the mitochondrial fusion-fission machinery, possibly acting on the mitochondrial fusion-associated protein mito-fusin-2 (Winter et al. 2015). Mitochondrial alterations in desmin KO cardiomyo-cytes were ameliorated by the overexpression of αB-crystallin, a chaperone of which mutations also cause heart failure (Diokmetzidou et al. 2016). This also improved cardiac functions which supports the idea that mitochondrial dysfunctions are a major feature of desmin loss in the heart.

In humans, the missense mutation Ile451Met located in the desmin tail domain causes a dilated cardiomyopathy. Expression of this mutant in transgenic mice revealed that mutant desmin lost its Z-disc localization but still associated with intercalated discs. The unaltered association with intercalated discs suggests that this association must take place through desmin (Mavroidis et al. 2008). In striated muscle, the plectin isoform 1d attaches desmin to Z-discs (Castanon et al. 2013) and interacts with the IF rod domain (Favre et al. 2011; Bouameur et al. 2014b). Interestingly, recruitment of desmin to cardiomyocyte intercalated discs can be inhibited by TNF-α-dependent mechanisms (Mann 2003; Panagopoulou et al. 2008). Cardiac-specific overexpression of secreted TNF-α in transgenic mice induced cardiomyopathy followed by caspase-6-dependent cleavage of desmin. This caused its aggregation and segregation from intercalated discs, with concomi-tant mislocalization of other proteins, such as desmoplakin and β-catenin, from those structures (Panagopoulou et al. 2008). In this setting, deletion of desmin pre-vented TNF-α-dependent desmoplakin and β-catenin detachment from intercalated discs. Conversely, desmoplakin and β-catenin localization were unaffected in des-min KO mice, showing that desmin is not necessary for their recruitment to interca-lated discs but rather for their maintenance, in analogy to keratins (Kroger et al. 2013). Mating of cardiac TNF-α overexpressing mice with animals expressing a caspase-resistant desmin variant revealed that desmin aggregates promote apoptosis and therefore actively participate in cardiac phenotypes caused by TNF-α overex-pression. Surprisingly, cardiac overexpression of TNF-α also attenuated desmin KO phenotypes with the notable absence of extensive fibrotic lesions and calcification, reduced mitochondrial abnormalities and a decrease of inflammation (Papathanasiou et al. 2015). Gene expression and functional analysis revealed that this improvement was due to a strong TNF-α-mediated increase of K8/K18 production by NF-κB-dependent mechanisms. This model showed that ectopically expressed K8/K18 can fulfill some of the functions of desmin in the heart intercalated discs but not at Z-discs from where they were absent. This model also proposes TNF-α, NF-κB or K8/K18 as potential targets in cardiac desminopathies. These data support the view that at least in the heart, K8 and K18 may be able to compensate for desmin.

In adult striated muscle, synemin, K8 and K19-containing filaments, in addition to desmin, localize at the Z-discs and link them to costameres at the sarcolemma (Goodall et al. 2012; Garcia-Pelagio et al. 2015). In analogy to cardiac muscle, a striated muscle contribution of keratin IF is supported by K19 KO mice. In these, costameric structures at tibialis anterior muscles are lost and specific tension was lower compared to desmin KO (Stone et al. 2007; Lovering et al. 2011). Both models

showed elevated serum levels of creatine kinase, indicating tissue lesions. Therefore, keratins and desmin contribute to costamere formation and/or maintenance and to myofibril alignment. Clearly, desmin performs a predominant role in this setting. As expected from single KOs, mice lacking both desmin and K19 showed reduced specific force of fast-twitch muscles with an increased susceptibility to injury (Lovering et al. 2011; Goodall et al. 2012; Shah et al. 2012). Of note, the distance between the most superficial myofibrils and the sarcolemma was significantly more prominent in K19 KO than desmin KO, indicating that in addition to desmin, the K19-containing keratin network of striated muscle is functionally relevant. Mitochondrial defects represent another facet in desmin-deficient muscle. Remarkably, overexpression of the anti-apoptotic bcl-2 gene in desmin null heart improved mitochondrial defects, hypertrophy and cardiac functions (Weisleder et al. 2004b). How loss of desmin affects mitochondria and thereby metabolism or apoptosis requires future analysis (Fountoulakis et al. 2005). Muscle anomalies were predominantly observed in slow twitch skeletal and in cardiac muscle displaying high mitochondrial activity and intensified with ageing and upon physical exercise. Further studies established that *in situ*, ADP-stimulated oxygen consumption in KO cardiac and soleus muscle was significantly reduced. This showed that desmin, similar to keratins (Kumar et al. 2015), affect mitochondrial localization, membrane composition and, ultimately, their activity. It is conceivable that desmin acts on mitochondria through the cytolinker plectin isoform 1b which is able to bind both mitochondria and desmin (Winter et al. 2008, 2015).

7.3.3   GFAP

GFAP is the major IF of mature astrocytes in the mammalian central nervous system and is also expressed in a subset of stem cells (Liem and Messing 2009; Brenner 2014; Cotrina et al. 2014; Hol and Pekny 2015). Several independent strains of KO mice have been generated, all of which are viable. Regarding their phenotype, contradictory observations on localized myelination defects, abnormalities of neuronal survival and susceptibility to central nervous system injury were reported (Gomi et al. 1995; Pekny et al. 1995; Liedtke et al. 1996; McCall et al. 1996; Hanbury et al. 2003). The potential reasons for the phenotypic differences among available GFAP KO mice were reviewed before (Pekny and Pekna 2004). In humans, missense GFAP mutations cause the progressive and fatal neurodegenerative Alexander disease (AxD), characterized by Rosenthal fibres containing cytoplasmic ubiquitinated protein aggregates of GFAP, vimentin, and small heat shock proteins in astrocytes. GFAP KI mice with GFAP-Arg76His and -Arg236His mutations develop Rosenthal fibres, the hallmark protein aggregates observed in astrocytes in AxD, in the hippocampus, corpus callosum, olfactory bulbs, subpial, and periventricular regions. GFAP mutant mice had a normal lifespan and showed no overt behavioural defects, but were more susceptible to kainate-induced seizures. Further increase in GFAP via crosses to GFAP transgenic animals led to a shift in GFAP solubility, an increased

stress response, and ultimately death (Hagemann et al. 2006). AxD mice presented with a deficit in adult neurogenesis due to abnormal proliferation and differentiation of neural progenitors in the hippocampus. This resulted in impairments in contextual learning and spatial memory (Hagemann et al. 2013). The severity of the phenotype was also dependent on levels of GFAP mutants (Jany et al. 2013). Moreover, AxD mouse models presented increased hyperactivity which did not increase seizure susceptibility after traumatic brain injury. In Rosenthal fibres, the chaperone αB-crystallin, is also present in the GFAP inclusions together with vimentin, plectin, ubiquitin and Hsp27. In an AxD mouse model in which GFAP was overexpressed, the severity of the phenotype correlated with expression levels of αB-crystallin suggesting its potential therapeutic targetting (Hagemann et al. 2009). This demonstrated the primary importance of GFAP aggregates rather than filaments loss in the AxD pathomechanism. In support, transactivation of the Gfap promoter is an early and sustained indicator of the disease process in the mouse. Furthermore, GFAP in cerebrospinal fluid serves as a potential biomarker that is comparable between mouse models and human patients (Jany et al. 2013).

During reactive gliosis, astrocytes upregulate the expression of GFAP, vimentin and nestin. GFAP and vimentin double KO mice are viable and develop mild phenotypes, primarily revealing a role of IFs in glial tissues during reactive gliosis and neuroregeneration. Illustrating this, GFAP/vimentin KO mice presented attenuated reactive gliosis and delayed axonal regeneration with altered re-innervated motor endplates on gastrocnemius muscle fibres after sciatic nerve crush (Pekny et al. 1999; Berg et al. 2013). Also, after cerebral or ischemic stroke those animals consistently presented larger infarcts and slightly reduced motor behaviour compared to WTs (de Pablo et al. 2013; Liu et al. 2014). A neuron-astrocyte co-culture model also confirmed that GFAP/vimentin astrocytes were less neuroprotective after oxygen–glucose deprivation with higher ROS accumulation than WTs probably explaining increased cell death. This also suggests that IFs could modulate oxidative stress response (de Pablo et al. 2013). In fact, GFAP or GFAP/vimentin double KO induced numerous changes in gene expression profiles of astrocytes, reflecting possible compensatory mechanisms or loss of IF functions. Those changes include genes involved in inflammatory response (Kamphuis et al. 2015). In fact, astrocytes are also antigen-presenting cells as they express cell-surface MHC class II molecules at their surface when stimulated by the cytokine IFN-γ (Shrikant and Benveniste 1996) and reactive astrocytes upregulate expression of IF proteins (Pekny and Nilsson 2005) suggesting that IFs could be involved in a stress response or signalling. Also, analysis of GFAP/vimentin double KO astrocytes revealed that IFs may participate in the vesicular transport of IFN-γ-induced MHC class II in late endosomes/lysosomes with inherent implications in neuroinflammation (Vardjan et al. 2012). Nevertheless, in some glial cells IFs may have other effects which rather limit normal neurogenesis. Astrocytes contribute to the control of adult neurogenesis through Notch signalling. This depends in part on GFAP and vimentin. In line, genetic deletion of both GFAP and vimentin showed an increased neurogenesis. Cultured astrocytes lacking GFAP and vimentin showed decreased Notch signalling and altered expression of Notch signalling pathway-related genes Dlk2, Notch1, and Sox2 (Pekny et al. 2014; Lebkuechner et al. 2015). *In vivo* and *in vitro*

analysis revealed that astrocytes that typically express those IF proteins attenuate normal and post-injury neuronal differentiation (Larsson et al. 2004; Cho et al. 2005; Widestrand et al. 2007; Wilhelmsson et al. 2012).

7.3.4 Syncoilin and peripherin

Syncoilin is expressed in two isoforms and is associated with the dystrophin-associated protein complex in striated muscle. In addition, syncoilin is expressed in both the central and peripheral nervous systems. Isoform Sync1 is dominant in the brain, whereas isoform Sync2 prevails in the spinal cord and sciatic nerve (Kemp et al. 2009; Clarke et al. 2010). In striated muscle, it interacts with desmin and is detected around cell nuclei, at the sarcolemma, at myotendinous and neuromuscular junctions. Syncoilin can also interact with and modulate the formation of peripherin filaments in neurons *in vitro* (Clarke et al. 2010). Syncoilin KO mice were reported to develop a subtle striated muscle phenotype with a slight reduction in isometric stress production and an increased susceptibility to muscle damage in response to an extended forced exercise regime (McCullagh et al. 2008; Zhang et al. 2008). In agreement with peripherin expression, syncoilin KO mice also showed a modest but significant reduced ability in accelerating treadmill and rotarod tests possibly due to motor axons defects (Clarke et al. 2010). Major mouse models for the analysis of type III IFs are listed in Table 7.3. Peripherin is predominantly expressed in the peripheral nervous system in adults (Hafidi 1998). Deletion of the peripherin gene in mice has so far not been connected to a severe pathology (Maison et al. 2016; Froud 2015; Larivière et al. 2002). Nevertheless, dysfunction of synaptic transmission between olivocochlear terminals and their peripheral targets could be detected in those animals. This suggested the conclusion that the type II cochlear ganglion neurons do not conduct efferent reflexes.

7.4 Analysis of Type IV Intermediate Filaments

7.4.1 Neuronal Type IV IFs

Neurofilaments are encoded by three genes and are obligate heteropolymers made of neurofilament light (NF-L), medium (NF-M), and heavy (NF-H) subunit proteins mostly diverging in tail domain sequences (Lee et al. 1993). They are expressed in neurons of the central and peripheral nervous systems and their ratio can vary according to the tissue. Early analysis of a spontaneous mutant of NF-L in quail revealed its prominent role in the filaments, compared to NF-M and -H, as the animals lost all axonal neurofilaments (Ohara et al. 1993). Neurofilaments are transported along axons in a complex manner (Brown 2003) and form with other cytoskeletal filaments a non-homogenous network. Several cytolinkers and motor proteins are able to interact directly or indirectly with them and regulate their

Table 7.3 Major animal models of type III intermediate filaments

Proteins	Genetic models	Lethality	Major phenotypes	References
Desmin	KO mice		Slight skeletal myopathy, dilated cardiomyopathy with cardiomyocyte death, fibrosis, calcification, mitochondrial localization and function defects, risk of fatal heart failure	Li et al. (1996), Milner et al. (1996), Milner et al. (1999), Boriek et al. (2001), Mavroidis and Capetanaki (2002), O'Neill et al. (2002), Fountoulakis et al. (2005), Ralston et al. (2006), and Goodall et al. (2012))
	Desmin/C5b receptor double KO		Improvement of the cardiac phenotype	Mavroidis and Capetanaki (2002) and Mavroidis et al. (2015)
	Desmin KO + cardiac spec. TNF-α expres.		Partial rescue of desmin KO caused cardiac fibrosis, calcification and mitochondrial dysfunction	Papathanasiou et al. (2015)
	Desmin KO + cardiac-spec. Transgenic bcl-2		Rescue of desmin KO cardiac phenotype	Weisleder et al. (2004b)
Desmin/K19	Double KO mice		Reduced specific force of fast-twitch muscles with an increased susceptibility to injury	Goodall et al. (2012)

GFAP				
	KO mice		Localized myelination defects, abnormalities of neuronal survival and susceptibility to central nervous system injury (age dependent defects and not observed in all models)	Gomi et al. (1995), Pekny et al. (1995), Liedtke et al. (1996), McCall et al. (1996) and Hanbury et al. (2003)
	GFAP$^{+/R76H}$ or GFAP$^{+/R236H}$ KI mice (equivalent to Alexander disease-associated mutation)		Rosenthal fibre formation, increased susceptibility to kainate-induced seizures, deficit in adult neurogenesis, impaired contextual learning and spatial memory	Hagemann et al. (2006), Hagemann et al. (2013), and Jany et al. (2013)
	hGFAP transgenic mice (overexpression)	Few days to more than a year depending on the transgene level of expression	Formation of Rosenthal fibres and white matter degeneration leading to neurological disorder with motor delays, spasticity, epileptic episodes and lethality	Messing et al. (1998) and Cotrina et al. (2014)
	hGFAP Tg αB-crystallin KO mice	~30 days of age (increased compared to hGFAP Tg)	Rosenthal fibres	Hagemann et al. (2009)
	hGFAP Tg αB-crystallin Tg mice	Increased survival depending on Tg αB-crystallin: hGFAP ratio	Rosenthal fibres, rescue of lethal phenotype in GFAP Tg mice (survival time ≥3 months), increased body weight compared with GFAP Tg littermates, reduced GFAP accumulation and stress response, αB-crystallin also rescues GFAPTg;GFAP$^{+/R236H}$ mice from lethal seizures and reduces Rosenthal fibres and GFAP in hippocampus	Hagemann et al. (2009)

(continued)

Table 7.2 (continued)

Proteins	Genetic models	Lethality	Major phenotypes	References
GFAP/Vimentin	double KO mice		Increased neurogenesis, impaired antigen presentation by astrocytes, attenuated reactive gliosis. After peripheral axotomy: delayed or reduced reinnervated motor endplates on muscle and delayed axonal regeneration. After oxygen-glucose deprivation in cultured cells: larger infarcts and higher ROS accumulation due to impaired elimination and increased cell death	Pekny et al. (1999), Larsson et al. (2004), Cho et al. (2005), Widestrand et al. (2007), Li et al. (2008), Vardjan et al. (2012), Wilhelmsson et al. (2012), Berg et al. (2013) and Lebkuechner et al. (2015)
Syncoilin/Peripherin	KO mice		Subtle striated muscle phenotype and increased susceptibility to muscle damage during extended exercise. Possible motor neuron defects. Peripherin: Dysfunction of synaptic transmission between olivocochlear terminals and their peripheral targets	McCullagh et al. (2008), Zhang et al. (2008), Clarke et al. (2010), Maison et al. (2016), Froud et al. (2015) and Larivière et al. (2002)
Vimentin	KO mice		No obvious phenotype in unchallenged mice. Impaired embryonic or adult skin wound healing, impaired leucocyte adhesion, migration and transmigration, increased bacterial killing during phagocytosis, resistance to extran sodium sulphate-induced acute colitis and acute lung injury, vimentin may act as negative regulator of NLRP3 inflammasome.	Colucci-Guyon et al. (1994), Eckes et al. (2000), Nieminen et al. (2006), Ivaska et al. (2007), Mor-Vaknin et al. (2013), dos Santos et al. (2015), dos Santos et al. (2015) and Cheng et al. (2016)
	Vim[R125C]		Aggregates, posterior cataract and Hsp70 upregulation, mutation analogous to K14 mutation in EBS	Bornheim et al. (2008) and Muller et al. (2009)
	Vim[HeadSA/SA] KI		Microophthalmia and cataract, delayed dermal wound healing with increased senescence (significant failure of cytokinesis)	Panagopoulou et al. (2008), Matsuyama et al. (2013), and Tanaka et al. (2015)

behaviour, including plectin (Errante et al. 1994; Rao et al. 2003), BPAGn/dystonin (Yang et al. 1996; Dalpe et al. 1998), gigaxonin (Bomont et al. 2000) and myosin Va (Rao et al. 2002a).

Several mouse models revealed that neurofilaments' predominant function lies in the control of axonal radial growth; however, genetic neurofilament ablation failed to cause obvious pathology (Lariviere and Julien 2004). All neurofilament KO mice present a variably reduced diameter of myelinated axons accompanied by a modest axonal loss in the cases of NF-L and -M KOs (Zhu et al. 1997, 1998; Elder et al. 1999a, 1999b; Jacomy et al. 1999). In NF-L-deficient animals, this was accompanied with slowed conduction velocities and delayed axon regeneration after injury (Sakaguchi et al. 1993; Zhu et al. 1997; Kriz et al. 2000). Accumulating atrophy of motor axons has been observed in NF-M KO mice (Elder et al. 1999a) and animals lacking both NF-M and -H did show similar phenotypes but with more pronounced anomalies leading to hind limb paralysis appearing with age. Phenotypes were not due to neuronal loss. Abnormalities in NF-L, NF-M and NF-M/-H KO mice correlated with reduced amounts of neurofilaments emphasizing their function in structural maintenance of neurons. In fact, a correct ratio of neurofilament subunits is critical as inferred by other mouse models with transgenic expression of the proteins. For instance, mice overexpressing hNF-H accumulated mainly perikaryal neurofilament inclusions causing progressive motor dysfunction (Cote et al. 1993). This phenomenon was balanced if both hNF-H and hNF-L are overexpressed (Meier et al. 1999). Similarly, mice overexpressing NF-L with either NF-M or NF-H had increased axonal diameter (Xu et al. 1996).

Also, a controlled transgenic expression of human NF-L in NF-L WT and KO mice provided insight into their stability and transport (Millecamps et al. 2007). Transgenic human NF-L had a half-life of 3 weeks in the brain, whereas in axons of the sciatic nerve it remained detectable for up to 8 months, indicating a significant stationary pool of neurofilament proteins. A highly stable pool of neurofilaments has also been reported in cultured mouse primary cortical neurons transfected with EGFP-NF-L (Yuan et al. 2009). Here, non-filamentous neurofilament subunits were slowly incorporated into the stable filaments. Along axons, their anterograde and retrograde transports are controlled by typical microtubule-dependent kinesin or dynein/dynactin transport machinery (Shah et al. 2000; Wang et al. 2000; Rao et al. 2002a; Francis et al. 2005). Correct neurofilament transport, involved in their assembly and stabilization, is mediated by NUDEL and myosin Va, proteins that interact with both NF-L or NF-H and dynein/dynactin complex important for controlling axonal transport of synaptic or membrane-bound vesicles but also organelles such as endoplasmic reticulum or mitochondria (Rao et al. 2002a; Nguyen et al. 2004). Moreover, myosin Va KO mice presented an increase of neurofilament content in line with the idea that neurofilament transport plays a role in their half-life (Rao et al. 2002a). Conversely, loss of NF-L resulted in a decreased amount, an altered transport and distribution of myosin Va in axons with consequences on organelle distribution including the endoplasmic reticulum (Rao et al. 2002a, 2011). This indicates that neurofilaments serve as substrates for myosin Va-dependent transport and that the phenotypes caused upon deletion of NF-L result from

transport defects. In addition, neurofilaments can directly interact with microtubules through an aminoterminal motif that is conserved among several IF proteins (Lepinoux-Chambaud and Eyer 2013). In transgenic mice, in which neurofilaments are withheld from the axonal compartment, axonal tubulin accumulation is normal but microtubules assemble in excessive numbers. Thus, axonal neurofilaments may modulate local microtubule assembly. This suggests a novel mechanism through which mutations in IF genes might contribute to pathogenesis in disease conditions (Bocquet et al. 2009; Lepinoux-Chambaud and Eyer 2013).

Neurofilaments are also synapse components where they are differently phosphorylated and behave differently than in the rest of the axons (Yuan et al. 2015). Their functional impact on neurotransmission was underlined by NF-M KO mice which presented amplified dopamine D1-receptor-mediated motor responses to cocaine, most probably because of an increased translocation of postsynaptic D1-receptors from endosomes to plasma membrane. Besides, the triple KO of α-internexin, NF-H and NF-L depresses hippocampal long-term potentiation induction. The fact that single KO mice for NF-H or of NF-L did not develop neurotransmission defects points to a specific role of NF-M, possibly through sequestration of postsynaptic receptors.

In human and mice, mutations of NF-L or the superoxide dismutase 1 (SOD1) genes cause amyotrophic lateral sclerosis, where abnormal expression of IF proteins could be seen in neurons (Rouleau et al. 1996). Transgenic mice expressing disease-causing mutations of SOD1 mated with different transgenic or KO mice for neurofilaments revealed that the latter do not participate in the pathogenesis initiated by SOD1 (Couillard-Despres et al. 1998, 2000; Kong et al. 1998). However, there were inverse correlations between the amounts of accumulated perikaryal neurofilaments and the extent of SOD1 mutant phenotype revealing protective effects of perikaryal neurofilaments. In this context, perikaryal neurofilaments colocalized with cdk5 and p25 kinases involved in the SOD1 mutant pathogenesis via toxic hyperphosphorylation of Tau protein notably (Nguyen et al. 2001). Presence of perikaryal neurofilaments also correlated with decreased Tau phosphorylation suggesting that accumulated perikaryal neurofilaments could act as scaffold or as reservoir of substrate for kinases which attenuates the phenotype.

NF-L mutations also cause Charcot-Marie-Tooth disease type 2E, another neurodegenerative disease with sensorimotor neuropathies slowly leading to progressive muscle weakness and atrophy (De Jonghe and Jordanova 1993). In humans, NF-L[P22S] mutation leads to filament aggregation and is pathogenic (Sasaki et al. 2006). Expression of an equivalent mutation in mice partially mimicked the human condition, in particular with regard to the extended duration of the hindlimb clasping response and gait anomalies, sensorimotor deficits in stationary beam, suspended bar tests and deficiency in the reversal phase of left-right discrimination learning in a water maze (Filali et al. 2011). Another disease-causing mutation, NF-L[N98S] has also been introduced in mice. These developed tremor and hindlimb clasping dominant phenotypes involving the cerebellum and the cerebral cortex and pons (Adebola et al. 2015). As predicted, the mutation caused protein aggregation and showed reduced amounts of neurofilaments and a reduced axonal diameter, in agreement

with observations in KO mice. Interestingly, a third NF-L disease-causing mutation, P8R, did not show any effect when introduced in mice even in homozygous lines (Adebola et al. 2015). Further studies on these animals may elucidate additional molecular pathomechanisms involved in Charcot-Marie-Tooth disease type 2E.

Correspondingly, KI mouse models with carboxyl-terminal tail-deletions of NF-M or both NF-M and NF-H also showed diminished radial axonal growth with altered neurofilament organization but with intact transport emphasizing the particular importance of the KSP motif in this domain (Rao et al. 2002b; Garcia et al. 2003; Yuan et al. 2006). Early studies have shown that carboxyl-terminal hyperphosphorylation of neurofilaments, notably in KSP, KXSP or KXXSP repeat motifs, was associated with neurofilament-dependent radial growth (de Waegh et al. 1992; Yin et al. 1998). Neurofilament phosphorylation was also associated with crossbridge formation to microtubules and an increased axonal diameter (Shea and Lee 2011). It is detected primarily in axons for NF-M, NF-H. In contrast to expectation, mice that express a KSP phosphorylation-deficient mutant of NF-M, showed normal axonal calibre and motor neuron conduction velocity. This established that proper phosphorylation of this domain is not required for radial axonal growth during myelination (Garcia et al. 2009).

7.4.2 Synemin

Synemin is mainly expressed in glia, neurons and striated muscle (Izmiryan et al. 2006; Mizuno et al. 2009; Prudner et al. 2014). The synemin gene encodes three isoforms (H or α, M or β and L) with different expression pattern and sizes that result from their carboxyl-terminal domains (Izmiryan et al. 2009). All synemin isoforms form obligate heteropolymers with desmin, vimentin, GFAP, peripherin or neurofilaments (Jing et al. 2007; Izmiryan et al. 2009; de Souza Martins et al. 2011). As an anchoring protein for protein kinase A (PKA), it provides temporal and spatial targetting of the enzyme. Two studies have analyzed the consequences of synemin deletion in mice. Both groups found that synemin KO mice are viable and develop a mild skeletal muscle phenotype. Tibialis anterior muscles show a significant decrease in mean fibre diameter, a decrease in twitch and tetanic force, and an increase in susceptibility to injury caused by lengthening contractions. Despite its distinct subcellular localization, the organization of proteins associated with the contractile apparatus and costameres was not significantly altered in KO mice. When exposed to exercise, the ability of KO mice to run uphill on a treadmill was similar to controls (Garcia-Pelagio et al. 2015). The second study reported similar findings and additionally investigated its signalling role. In the absence of synemin, decreased myostatin atrogin and increased follistatin expression were found. In line with its function toward PKA, the PKA RIIα subunit, p70S6K and CREB1 were increased in mutant mice. To investigate functional consequences of synemin ablation during muscle regeneration, the behaviour of satellite cells, which express it, was studied. In the setting of muscle overload, the number of Pax7-positive satellite

cells increased in control, but not in synemin KO mice. These findings indicate that the absence of synemin might affect the balance between self-renewal and differentiation of muscle satellite cells (Li et al. 2014). It has also been shown that those animals have osteopenic bone indicated by reduced bone formation markers and altered osteoblast proliferation and differentiation (Moorer et al. 2016). Remarkably, primary osteoblasts isolated from synemin-null mice demonstrate markedly enhanced osteogenic capacity with a concomitant reduction in cyclin D1 mRNA expression, which may explain the loss of osteoblast number observed in vivo.

7.4.3   Nestin

The type IV IF protein nestin is widely expressed in proliferating neuroepithelial progenitor cells and in several adult tissues. Astrocytes, for example, express GFAP, vimentin, and nestin. In addition to full-length nestin, two other isoforms (nestin-short and Nes-SΔ107-254) are known. These are exclusively expressed in the otherwise nestin-negative dorsal root ganglia (Lindqvist et al. 2016). Studies of IF networks in astrocytes from GFAP and vimentin KO mice showed that nestin cannot form filaments on its own and that it has to associate with other IFs to be detected in a filamentous stage in astrocytes (Eliasson et al. 1999). Similar to most IF proteins, nestin is highly phosphorylated. Among the best known regulatory enzymes, cdc2, cdk5 and p35, numerous studies have led to the hypothesis that nestin serves as a scaffold at the neuromuscular synapse that controls the activity of these proteins locally (Sahlgren et al. 2001, 2003). The nestin KO mice, generated by two groups, developed distinct and contradictory phenotypes. Mice in which the first exon of the nestin gene was deleted, suffered from embryonic lethality at ≈E8.5 (Park et al. 2010). It is not clear whether the nestin gene in these mice was completely deleted. The neuroepithelium of the developing neural tube exhibited significantly fewer neural stem cells and much higher levels of apoptosis. In a second study, the first exon was replaced with a selectable marker gene, leading to mice that were viable. This suggested that nestin deficiency is compatible with normal development of the CNS. Adult mice, however, showed impaired motor coordination. At the cellular level, the number of acetylcholine receptor clusters, the nerve length, and the end-plate bandwidth were significantly increased in the neuromuscular junction area of nestin KO mice. This was similar to the phenotype described for deficiency of Cdk5 KO mice. The finding that nestin deficiency rescued the maintenance of acetylcholine receptor clusters in the absence of agrin, similar to mice lacking both Cdk5 and agrin, suggested that the phenotype resulting from nestin's absence is the consequence of aberrant Cdk5 activity (Mohseni et al. 2011). This provides strong evidence for a scaffolding role of nestin in the localized control of Cdk5-mediated maintenance and dispersion of AChR clusters (Mohseni et al. 2011).

7.5 Type VI Intermediate Filaments

7.5.1 *Filensin (BFSP1) and Phakinin (BFSP2)*

Beaded Filament Structural Protein 1 (also named filensin or CP115), together with Beaded Filament Structural Protein 2 (also named phakinin or CP49) form the specialized IF system of so-called beaded filaments in terminally differentiated lens fibre cells (Fitzgerald et al. 2016). Mutations in these genes have been associated with juvenile-onset, progressive cataracts (Song et al. 2009). Unexpectedly, deletion of the phakinin gene in mice caused no cataract, despite concomitant down regulation of its partner filensin. However, light scattering and lens optical properties were sub-optimal in phakinin KO lenses compared to litter mate controls. Moreover, dramatic changes in plasma membrane organization of fibre cells were noted (Sandilands et al. 2003). In a subsequent study, residual vimentin IF were found, which may explain the lack of cataracts (Sandilands et al. 2004). Upon deletion of filensin, very similar findings were reported, namely destabilization of phakinin, evidence of light scattering, but lack of cataracts (Alizadeh et al. 2003). Heterozygous animals exhibited an intermediate phenotype, showing a reduction in filensin transcript and moderate light-scattering at 5 months. Although it is not yet fully resolved how lens IF proteins are anchored to lens fibre membranes (Song et al. 2009), aquaporin members participate in this. In particular, AQP0 localizes at the fibre cell membrane, and could provide anchorage for cytoskeletal structures, including lens IF proteins. Together, they may confer fibre cell shape, architecture and integrity (Kumari and Varadaraj 2014). As in other tissues, phakinin and filensin, together with vimentin (Bornheim et al. 2008; Muller et al. 2009) achieve this by supporting membrane adhesion structures and by interaction with cytoplasmic-associated proteins. Collectively, and in analogy to other IF, mutations in lens IF proteins can cause or predispose to disease. In view of the endurance of fibre cells (a lifetime in mammals), the maintenance of the optical properties of the lens represents a major IF function. Major mouse models for the analysis of type IV and type VI IFs are listed in Table 7.4.

7.6 *Caenorhabditis elegans* cytoplasmic IFs

The genome of *C. elegans* comprises 11 cytoplasmic IF genes that are differentially expressed in specific combinations, similar to mammalian IF and a single nuclear lamin gene (Karabinos et al. 2001). RNAi studies have demonstrated that two out of 11 *C. elegans* IF genes are essential for the development of larval epidermis (Hapiak et al. 2003; Karabinos et al. 2004; Woo et al. 2004) and three for intestine function and feeding (Karabinos et al. 2001, 2003; Bossinger et al. 2004; Karabinos et al. 2004). During development, some isotypes are also able to regulate the width of the lumen (Karabinos et al. 2001; Kolotuev et al. 2013). Exploiting the power of

Table 7.4 Major animal models of type IV and VI intermediate filaments

Proteins	Type	Genetic models	Major phenotypes	References
Nestin	IV	KO mice	Impaired motor coordination, increased nerve length and endplate bandwidth, increased number of acetylcholine receptor clusters in neuromuscular junction	Park et al. (2010) and Mohseni et al. (2011)
			Independent KO: midgestation mortality with full penetrance, reduction of neuronal stem cells in developing neural tube, highly elevated levels of apoptosis in cultured neuronal stem cells derived from nestin KO mice	
NF-L	IV	Spontaneous mutant quail	Loss of all axonal neurofilaments	Ohara et al. (1993)
	IV	KO mice	Smaller calibre myelinated axons with reduced conduction velocities, slightly impaired motor coordination and increased metabolism	Sakaguchi et al. (1993), Zhu et al. (1997), Kriz et al. (2000), Dubois et al. (2005a), and Dubois et al. (2005b)
	IV	N98S KI mice	Charcot-Marie-Tooth Type 2E phenotype; mutant protein-containing cytoplasmic aggregates and reduced amount of neurofilaments; reduced axonal diameter; disorganized processes in the cerebellum and abnormal processes in the cerebral cortex and pons	Adebola et al. (2015)
	IV	P22S KI mice	Partial Charcot-Marie-Tooth Type 2E phenotype	Filali et al. (2011)
	IV	NFL Tg overexpression mice	Neurofilament inclusions and increased axonal diameter	Xu et al. (1996) and Meier et al. (1999)
NF-M	IV	KO mice	Smaller calibre myelinated axons; amplified dopamine D1-receptor-mediated motor responses to cocaine	Elder et al. (1999a) and Yuan et al. (2015)
	IV	NF-MS→A phospho-deficient tail KI mice	Normal axonal calibre and motor neuron conduction velocity	Garcia et al. (2009)

(continued)

Table 7.4 (continued)

Proteins	Type	Genetic models	Major phenotypes	References
NF-H	IV	KO mice	Very modest reduction in the diameter of large ventral roots axons	Zhu et al. (1998)
	IV	Human or mouse NF-H transgenic mice	Substantial neurofilament inclusions and increased axonal diameter. Some motor dysfunction with human NF-H overexpression (fine tremors, abnormal limb flexions and signs of weakness)	Cote et al. (1993) and Marszalek et al. (1996)
NF-M, NF-H	IV	Double KO mice	Significant reduction of large ventral root axon diameter	Elder et al. (1999b) and Jacomy et al. (1999)
	IV	NFM Δtail or NF-(M-H) Δtail KI mice	Diminished radial axonal growth with altered neurofilament organization but with intact transport	Rao et al. (2002b) and Garcia et al. (2003)
Synemin	IV	KO mice	Reduced bone mass, mild decrease of skeletal muscle size with smaller striated muscle fibres and increased sarcolemmal deformability, susceptibility to injury during lengthening contractions	Garcia-Pelagio et al. (2015) and Moorer et al. (2016)
Filensin	VI	KO mice	No cataract, downregulation of phakinin, deteriorated light scatter and lens optical properties	Alizadeh et al. (2003)
Phakinin	VI	KO mice	No cataract, downregulation of filensin, deteriorated light scatter and lens optical properties	Sandilands et al. (2003) and Sandilands et al. (2004)

C. elegans genetics, the intestinal IF organizer IFO-1 was shown to colocalize with IF proteins and F-actin in the intestine. Here, it anchors different cytoskeletal components to the *C. elegans* intestinal terminal web (Carberry et al. 2012). Moreover, *C. elegans* tissues represent an accessible model system for monitoring tissue mechanics at high precision, in order to dissect IF function (Jahnel et al. 2016). Related studies demonstrated how IF proteins are organized by mechanical force acting on *C. elegans* hemidesmosomes, which link body-wall muscles and the epidermis. Upon hemidesmosome stimulation, p21-activated kinase mediates phosphorylation of intermediate filament proteins to reorganize them according to mechanical load (Zhang et al. 2011).

7.7   Conclusion

The comprehensive analysis of IF genes in transgenic mice, in combination with studies of human disease, has identified some IF proteins as vital determinants of cell and tissue architecture, adhesion and resilience to mechanical stress. Loss of function studies have also demonstrated that tissue morphogenesis and regeneration are not fundamentally disrupted by the absence of IF. These comprise epidermal keratins and desmin in mammals and corresponding IF genes in *C. elegans*. In these tissues that are exposed to considerable mechanical strain, KO and missense mutations cause similar, albeit sometimes distinct, defects supporting a major role of epidermal keratins and desmin as true cytoskeletal proteins. Loss of function studies for other IF genes revealed that they are not vital for the above functions but serve as protein scaffolds that protect cells and tissues under conditions of stress in a context-dependent manner. Overall, transgenic studies provide some support to the concept of redundancy among IF proteins, provided substitution experiments are performed such that expression levels are comparable. A subset of IF proteins, exemplified by keratins K1, K16, K17, desmin and nestin, seems to stand out by their role in cell signalling in a context-dependent manner. In support of their scaffolding or regulatory role are observations that the pathology resulting from their absence can be genetically rescued by transgenic expression of non-IF components, e.g., IL-18 or Bcl-2 (Weisleder et al. 2004b; Roth et al. 2012).

These observations raise a number of topics that in the near future can be addressed in a better way than before, owing to recent advances:

1. What are the isotype-specific functions of IF proteins and how are they encoded? This requires identification of signalling pathways that regulate IF posttranslational modifications in a context-dependent manner and of associated proteins. Here, model organisms, including Drosphila and *C. elegans* could be highly valuable. In particular, well-designed replacement experiments using Crispr/Cas are in great need

2. Do the various IFopathies share similar pathomechanisms and, if so, what is their molecular basis? The notion that expression of a mutant keratin in the fly, which by itself lacks cytoplasmic IF, caused a dominant-negative phenotype related to the human condition, suggests that Drosophila is valuable for the identification of pathways altered by mutant IF proteins and for development of therapies (Bohnekamp et al. 2015, 2016). Resolving disease mechanisms will require advanced animal models, in combination with organotypic cell culture models and medicinal chemistry approaches

3. To what extent do IF proteins contribute to tumourigenesis and metastasis?

Surprisingly, very few studies using transgenic animal models have been performed. In particular, keratins have emerged as IF proteins which significantly contribute to cell mechanics and modulation of immune responses (Depianto et al. 2010; Roth et al. 2012; Ramms et al. 2013; Seltmann et al. 2013a; Hobbs et al. 2015). Given that these properties are major determinants of malignancy, there is an

urgent need to address the mechanisms by which keratins and additional IF proteins perform their scaffolding and regulatory functions, both *in vivo* and *in vitro*.

References

Adebola AA, Di Castri T, He CZ, Salvatierra LA, Zhao J, Brown K, Lin CS, Worman HJ, Liem RK (2015) Neurofilament light polypeptide gene N98S mutation in mice leads to neurofilament network abnormalities and a Charcot-Marie-Tooth type 2E phenotype. Hum Mol Genet 24(8):2163–2174. doi:10.1093/hmg/ddu736

Alam CM, Silvander JS, Daniel EN, Tao GZ, Kvarnstrom SM, Alam P, Omary MB, Hanninen A, Toivola DM (2013) Keratin 8 modulates beta-cell stress responses and normoglycaemia. J Cell Sci 126(Pt 24):5635–5644. doi:10.1242/jcs.132795

Alizadeh A, Clark J, Seeberger T, Hess J, Blankenship T, FitzGerald PG (2003) Targeted deletion of the lens fiber cell-specific intermediate filament protein filensin. Invest Ophthalmol Vis Sci 44(12):5252–5258

Allen EH, Courtney DG, Atkinson SD, Moore JE, Mairs L, Toftgaard Poulsen E, Schiroli D, Maurizi E, Cole C, Hickerson RP, James J, Murgatroyd H, Smith FJ, MacEwen C, Enghild JJ, Nesbit MA, Leslie Pedrioli DM, McLean WH, Moore CB (2016) Keratin 12 missense mutation induces the unfolded protein response and apoptosis in Meesmann epithelial corneal dystrophy. Hum Mol Genet. doi:10.1093/hmg/ddw001

Alvarado DM, Coulombe PA (2014) Directed expression of a chimeric type II keratin partially rescues keratin 5-null mice. J Biol Chem 289(28):19435–19447. doi:10.1074/jbc.M114.553867

Ambatipudi S, Bhosale PG, Heath E, Pandey M, Kumar G, Kane S, Patil A, Maru GB, Desai RS, Watt FM, Mahimkar MB (2013) Downregulation of keratin 76 expression during oral carcinogenesis of human, hamster and mouse. PLoS One 8(7):e70688. doi:10.1371/journal.pone.0070688

Ameen NA, Figueroa Y, Salas PJ (2001) Anomalous apical plasma membrane phenotype in CK8-deficient mice indicates a novel role for intermediate filaments in the polarization of simple epithelia. J Cell Sci 114(Pt 3):563–575

Arin MJ, Roop DR (2004) Inducible mouse models for inherited skin diseases: implications for skin gene therapy. Cells Tissues Organs 177(3):160–168. doi:10.1159/000079990

Arin MJ, Oji V, Emmert S, Hausser I, Traupe H, Krieg T, Grimberg G (2011) Expanding the keratin mutation database: novel and recurrent mutations and genotype-phenotype correlations in 28 patients with epidermolytic ichthyosis. Br J Dermatol 164(2):442–447. doi:10.1111/j.1365-2133.2010.10096.x

Asghar MN, Silvander JS, Helenius TO, Lahdeniemi IA, Alam C, Fortelius LE, Holmsten RO, Toivola DM (2015) The amount of keratins matters for stress protection of the colonic epithelium. PLoS One 10(5):e0127436. doi:10.1371/journal.pone.0127436

Bar J, Kumar V, Roth W, Schwarz N, Richter M, Leube RE, Magin TM (2014) Skin fragility and impaired desmosomal adhesion in mice lacking all keratins. J Invest Dermatol 134(4):1012–1022. doi:10.1038/jid.2013.416

Baribault H, Price J, Miyai K, Oshima RG (1993) Mid-gestational lethality in mice lacking keratin 8. Genes Dev 7(7A):1191–1202

Baribault H, Penner J, Iozzo RV, Wilson-Heiner M (1994) Colorectal hyperplasia and inflammation in keratin 8-deficient FVB/N mice. Genes Dev 8(24):2964–2973

Berg A, Zelano J, Pekna M, Wilhelmsson U, Pekny M, Cullheim S (2013) Axonal regeneration after sciatic nerve lesion is delayed but complete in GFAP- and vimentin-deficient mice. PLoS One 8(11):e79395. doi:10.1371/journal.pone.0079395

Bianchi N, Depianto D, McGowan K, Gu C, Coulombe PA (2005) Exploiting the keratin 17 gene promoter to visualize live cells in epithelial appendages of mice. Mol Cell Biol 25(16):7249–7259. doi:10.1128/mcb.25.16.7249-7259.2005

Block J, Schroeder V, Pawelzyk P, Willenbacher N, Koster S (2015) Physical properties of cytoplasmic intermediate filaments. Biochim Biophys Acta 1853(11 Pt B):3053–3064. doi:10.1016/j.bbamcr.2015.05.009

Bocquet A, Berges R, Frank R, Robert P, Peterson AC, Eyer J (2009) Neurofilaments bind tubulin and modulate its polymerization. J Neurosci 29(35):11043–11054. doi:10.1523/JNEUROSCI.1924-09.2009

Bohnekamp J, Cryderman DE, Paululat A, Baccam GC, Wallrath LL, Magin TM (2015) A Drosophila model of epidermolysis Bullosa Simplex. J Invest Dermatol 135(8):2031–2039. doi:10.1038/jid.2015.129

Bomont P (2016) Degradation of the intermediate filament family by Gigaxonin. Methods Enzymol 569:215–231. doi:10.1016/bs.mie.2015.07.009

Bomont P, Cavalier L, Blondeau F, Ben Hamida C, Belal S, Tazir M, Demir E, Topaloglu H, Korinthenberg R, Tuysuz B, Landrieu P, Hentati F, Koenig M (2000) The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. Nat Genet 26(3):370–374. doi:10.1038/81701

Boriek AM, Capetanaki Y, Hwang W, Officer T, Badshah M, Rodarte J, Tidball JG (2001) Desmin integrates the three-dimensional mechanical properties of muscles. Am J Physiol Cell Physiol 280(1):C46–C52

Bornheim R, Muller M, Reuter U, Herrmann H, Bussow H, Magin TM (2008) A dominant vimentin mutant upregulates Hsp70 and the activity of the ubiquitin-proteasome system, and causes posterior cataracts in transgenic mice. J Cell Sci 121(Pt 22):3737–3746. doi:10.1242/jcs.030312

Bose A, Teh MT, Mackenzie IC, Waseem A (2013) Keratin k15 as a biomarker of epidermal stem cells. Int J Mol Sci 14(10):19385–19398. doi:10.3390/ijms141019385

Bossinger O, Fukushige T, Claeys M, Borgonie G, McGhee JD (2004) The apical disposition of the *Caenorhabditis elegans* intestinal terminal web is maintained by LET-413. Dev Biol 268(2):448–456. doi:10.1016/j.ydbio.2004.01.003

Bouameur JE, Favre B, Borradori L (2014a) Plakins, a versatile family of cytolinkers: roles in skin integrity and in human diseases. J Invest Dermatol 134(4):885–894. doi:10.1038/jid.2013.498

Bouameur JE, Favre B, Fontao L, Lingasamy P, Begre N, Borradori L (2014b) Interaction of plectin with keratins 5 and 14: dependence on several plectin domains and keratin quaternary structure. J Invest Dermatol 134(11):2776–2783. doi:10.1038/jid.2014.255

Brenner M (2014) Role of GFAP in CNS injuries. Neurosci Lett 565:7–13. doi:10.1016/j.neulet.2014.01.055

Brown A (2003) Live-cell imaging of slow axonal transport in cultured neurons. Methods Cell Biol 71:305–323

Byrne C, Tainsky M, Fuchs E (1994) Programming gene expression in developing epidermis. Development 120(9):2369–2383

Candi E, Schmidt R, Melino G (2005) The cornified envelope: a model of cell death in the skin. Nat Rev Mol Cell Biol 6(4):328–340. doi:10.1038/nrm1619

Capetanaki Y, Bloch RJ, Kouloumenta A, Mavroidis M, Psarras S (2007) Muscle intermediate filaments and their links to membranes and membranous organelles. Exp Cell Res 313(10):2063–2076. doi:10.1016/j.yexcr.2007.03.033

Carberry K, Wiesenfahrt T, Geisler F, Stocker S, Gerhardus H, Uberbach D, Davis W, Jorgensen E, Leube RE, Bossinger O (2012) The novel intestinal filament organizer IFO-1 contributes to epithelial integrity in concert with ERM-1 and DLG-1. Development 139(10):1851–1862. doi:10.1242/dev.075788

Castanon MJ, Walko G, Winter L, Wiche G (2013) Plectin-intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. Histochem Cell Biol 140(1):33–53. doi:10.1007/s00418-013-1102-0

Caulin C, Ware CF, Magin TM, Oshima RG (2000) Keratin-dependent, epithelial resistance to tumor necrosis factor-induced apoptosis. J Cell Biol 149(1):17–22

Chen J, Cheng X, Merched-Sauvage M, Caulin C, Roop DR, Koch PJ (2006) An unexpected role for keratin 10 end domains in susceptibility to skin cancer. J Cell Sci 119(Pt 24):5067–5076. doi:10.1242/jcs.03298

Chen J, Jaeger K, Den Z, Koch PJ, Sundberg JP, Roop DR (2008) Mice expressing a mutant Krt75 (K6hf) allele develop hair and nail defects resembling pachyonychia congenita. J Invest Dermatol 128(2):270–279. doi:10.1038/sj.jid.5701038

Chen Y, Guldiken N, Spurny M, Mohammed HH, Haybaeck J, Pollheimer MJ, Fickert P, Gassler N, Jeon MK, Trautwein C, Strnad P (2015) Loss of keratin 19 favours the development of cholestatic liver disease through decreased ductular reaction. J Pathol 237(3):343–354. doi:10.1002/path.4580

Chernyatina AA, Guzenko D, Strelkov SV (2015) Intermediate filament structure: the bottom-up approach. Curr Opin Cell Biol 32:65–72. doi:10.1016/j.ceb.2014.12.007

Cheung KJ, Gabrielson E, Werb Z, Ewald AJ (2013) Collective invasion in breast cancer requires a conserved basal epithelial program. Cell 155(7):1639–1651. doi:10.1016/j.cell.2013.11.029

Cho KS, Yang L, Lu B, Feng Ma H, Huang X, Pekny M, Chen DF (2005) Re-establishing the regenerative potential of central nervous system axons in postnatal mice. J Cell Sci 118(Pt 5):863–872. doi:10.1242/jcs.01658

Chou CF, Smith AJ, Omary MB (1992) Characterization and dynamics of O-linked glycosylation of human cytokeratin 8 and 18. J Biol Chem 267(6):3901–3906

Chung BM, Arutyunov A, Ilagan E, Yao N, Wills-Karp M, Coulombe PA (2015) Regulation of C-X-C chemokine gene expression by keratin 17 and hnRNP K in skin tumor keratinocytes. J Cell Biol 208(5):613–627. doi:10.1083/jcb.201408026

Clarke WT, Edwards B, McCullagh KJ, Kemp MW, Moorwood C, Sherman DL, Burgess M, Davies KE (2010) Syncoilin modulates peripherin filament networks and is necessary for large-calibre motor neurons. J Cell Sci 123(Pt 15):2543–2552. doi:10.1242/jcs.059113

Cleveland DW, Yamanaka K, Bomont P (2009) Gigaxonin controls vimentin organization through a tubulin chaperone-independent pathway. Hum Mol Genet 18(8):1384–1394. doi:10.1093/hmg/ddp044

Colas J, Faure G, Saussereau E, Trudel S, Rabeh WM, Bitam S, Guerrera IC, Fritsch J, Sermet-Gaudelus I, Davezac N, Brouillard F, Lukacs GL, Herrmann H, Ollero M, Edelman A (2012) Disruption of cytokeratin-8 interaction with F508del-CFTR corrects its functional defect. Hum Mol Genet 21(3):623–634. doi:10.1093/hmg/ddr496

Colucci-Guyon E, Portier MM, Dunia I, Paulin D, Pournin S, Babinet C (1994) Mice lacking vimentin develop and reproduce without an obvious phenotype. Cell 79(4):679–694

Cote F, Collard JF, Julien JP (1993) Progressive neuronopathy in transgenic mice expressing the human neurofilament heavy gene: a mouse model of amyotrophic lateral sclerosis. Cell 73(1):35–46

Cotrina ML, Chen M, Han X, Iliff J, Ren Z, Sun W, Hagemann T, Goldman J, Messing A, Nedergaard M (2014) Effects of traumatic brain injury on reactive astrogliosis and seizures in mouse models of Alexander disease. Brain Res 1582:211–219. doi:10.1016/j.brainres.2014.07.029

Cotsarelis G (2006) Epithelial stem cells: a folliculocentric view. J Invest Dermatol 126(7):1459–1468. doi:10.1038/sj.jid.5700376

Couillard-Despres S, Zhu Q, Wong PC, Price DL, Cleveland DW, Julien JP (1998) Protective effect of neurofilament heavy gene overexpression in motor neuron disease induced by mutant superoxide dismutase. Proc Natl Acad Sci U S A 95(16):9626–9630

Couillard-Despres S, Meier J, Julien JP (2000) Extra axonal neurofilaments do not exacerbate disease caused by mutant Cu,Zn superoxide dismutase. Neurobiol Dis 7(4):462–470. doi:10.1006/nbdi.2000.0296

Coulombe PA, Lee CH (2012) Defining keratin protein function in skin epithelia: epidermolysis bullosa simplex and its aftermath. J Invest Dermatol 132(3 Pt 2):763–775. doi:10.1038/jid.2011.450

Coulombe PA, Kerns ML, Fuchs E (2009) Epidermolysis bullosa simplex: a paradigm for disorders of tissue fragility. J Clin Invest 119(7):1784–1793. doi:10.1172/JCI38177

Dalpe G, Leclerc N, Vallee A, Messer A, Mathieu M, De Repentigny Y, Kothary R (1998) Dystonin is essential for maintaining neuronal cytoskeleton organization. Mol Cell Neurosci 10(5–6):243–257. doi:10.1006/mcne.1997.0660

Davis BK, Wen H, Ting JP (2011) The inflammasome NLRs in immunity, inflammation, and associated diseases. Annu Rev Immunol 29:707–735. doi:10.1146/annurev-immunol-031210-101405

De Berker D, Wojnarowska F, Sviland L, Westgate GE, Dawber RP, Leigh IM (2000) Keratin expression in the normal nail unit: markers of regional differentiation. Br J Dermatol 142(1):89–96

De Jonghe P, Jordanova AK (1993) Charcot-Marie-Tooth neuropathy type 2E/1F. In: Pagon RA, Adam MP, Ardinger HH et al. (eds) GeneReviews(R). University of Seattle, Seattle

de Pablo Y, Nilsson M, Pekna M, Pekny M (2013) Intermediate filaments are important for astrocyte response to oxidative stress induced by oxygen-glucose deprivation and reperfusion. Histochem Cell Biol 140(1):81–91. doi:10.1007/s00418-013-1110-0

de Souza Martins SC, Agbulut O, Diguet N, Larcher JC, Paulsen BS, Rehen SK, Moura-Neto V, Paulin D, Li Z, Xue Z (2011) Dynamic expression of synemin isoforms in mouse embryonic stem cells and neural derivatives. BMC Cell Biol 12:51. doi:10.1186/1471-2121-12-51

de Waegh SM, Lee VM, Brady ST (1992) Local modulation of neurofilament phosphorylation, axonal caliber, and slow axonal transport by myelinating Schwann cells. Cell 68(3):451–463

Depianto D, Kerns ML, Dlugosz AA, Coulombe PA (2010) Keratin 17 promotes epithelial proliferation and tumor growth by polarizing the immune response in skin. Nat Genet 42(10):910–914. doi:10.1038/ng.665

Dequen F, Bomont P, Gowing G, Cleveland DW, Julien JP (2008) Modest loss of peripheral axons, muscle atrophy and formation of brain inclusions in mice with targeted deletion of gigaxonin exon 1. J Neurochem 107(1):253–264. doi:10.1111/j.1471-4159.2008.05601.x

DiTommaso T, Cottle DL, Pearson HB, Schluter H, Kaur P, Humbert PO, Smyth IM (2014) Keratin 76 is required for tight junction function and maintenance of the skin barrier. PLoS Genet 10(10):e1004706. doi:10.1371/journal.pgen.1004706

dos Santos G, Rogel MR, Baker MA, Troken JR, Urich D, Morales-Nebreda L, Sennello JA, Kutuzov MA, Sitikov A, Davis JM, Lam AP, Cheresh P, Kamp D, Shumaker DK, Budinger GR, Ridge KM (2015) Vimentin regulates activation of the NLRP3 inflammasome. Nat Commun 6:6574. doi:10.1038/ncomms7574

Duan Y, Sun Y, Zhang F, Zhang WK, Wang D, Wang Y, Cao X, Hu W, Xie C, Cuppoletti J, Magin TM, Wang H, Wu Z, Li N, Huang P (2012) Keratin K18 increases cystic fibrosis transmembrane conductance regulator (CFTR) surface expression by binding to its C-terminal hydrophobic patch. J Biol Chem 287(48):40547–40559. doi:10.1074/jbc.M112.403584

Dubois M, Lalonde R, Julien JP, Strazielle C (2005a) Mice with the deleted neurofilament of low-molecular-weight (Nefl) gene: 1. Effects on regional brain metabolism. J Neurosci Res 80(6):741–750. doi:10.1002/jnr.20449

Dubois M, Strazielle C, Julien JP, Lalonde R (2005b) Mice with the deleted neurofilament of low molecular weight (Nefl) gene: 2. Effects on motor functions and spatial orientation. J Neurosci Res 80(6):751–758. doi:10.1002/jnr.20493

Eckes B, Colucci-Guyon E, Smola H, Nodder S, Babinet C, Krieg T, Martin P (2000) Impaired wound healing in embryonic and adult mice lacking vimentin. J Cell Sci 113(Pt 13):2455–2462

Elder GA, Friedrich VL Jr, Margita A, Lazzarini RA (1999a) Age-related atrophy of motor axons in mice deficient in the mid-sized neurofilament subunit. J Cell Biol 146(1):181–192

Elder GA, Friedrich VL Jr, Pereira D, Tu PH, Zhang B, Lee VM, Lazzarini RA (1999b) Mice with disrupted midsized and heavy neurofilament genes lack axonal neurofilaments but have unaltered numbers of axonal microtubules. J Neurosci Res 57(1):23–32

Eliasson C, Sahlgren C, Berthold CH, Stakeberg J, Celis JE, Betsholtz C, Eriksson JE, Pekny M (1999) Intermediate filament protein partnership in astrocytes. J Biol Chem 274(34):23996–24006

Eriksson JE, Brautigan DL, Vallee R, Olmsted J, Fujiki H, Goldman RD (1992) Cytoskeletal integrity in interphase cells requires protein phosphatase activity. Proc Natl Acad Sci U S A 89(22):11093–11097

Eriksson JE, He T, Trejo-Skalli AV, Harmala-Brasken AS, Hellman J, Chou YH, Goldman RD (2004) Specific in vivo phosphorylation sites determine the assembly dynamics of vimentin intermediate filaments. J Cell Sci 117(Pt 6):919–932. doi:10.1242/jcs.00906

Eriksson JE, Dechat T, Grin B, Helfand B, Mendez M, Pallari HM, Goldman RD (2009) Introducing intermediate filaments: from discovery to disease. J Clin Invest 119(7):1763–1771. doi:10.1172/jci38339

Errante LD, Wiche G, Shaw G (1994) Distribution of plectin, an intermediate filament-associated protein, in the adult rat central nervous system. J Neurosci Res 37(4):515–528. doi:10.1002/jnr.490370411

Favre B, Schneider Y, Lingasamy P, Bouameur JE, Begre N, Gontier Y, Steiner-Champliaud MF, Frias MA, Borradori L, Fontao L (2011) Plectin interacts with the rod domain of type III intermediate filament proteins desmin and vimentin. Eur J Cell Biol 90(5):390–400. doi:10.1016/j.ejcb.2010.11.013

Feng X, Coulombe PA (2015) A role for disulfide bonding in keratin intermediate filament organization and dynamics in skin keratinocytes. J Cell Biol 209(1):59–72. doi:10.1083/jcb.201408079

Filali M, Dequen F, Lalonde R, Julien JP (2011) Sensorimotor and cognitive function of a NEFL(P22S) mutant model of Charcot-Marie-Tooth disease type 2E. Behav Brain Res 219(2):175–180. doi:10.1016/j.bbr.2010.12.022

Fischer H, Langbein L, Reichelt J, Praetzel-Wunder S, Buchberger M, Ghannadan M, Tschachler E, Eckhart L (2014) Loss of keratin K2 expression causes aberrant aggregation of K10, hyperkeratosis, and inflammation. J Invest Dermatol 134(10):2579–2588. doi:10.1038/jid.2014.197

Fischer H, Langbein L, Reichelt J, Buchberger M, Tschachler E, Eckhart L (2015) Keratins K2 and K10 are essential for the epidermal integrity of plantar skin. J Dermatol Sci. doi:10.1016/j.jdermsci.2015.10.008

Fountoulakis M, Soumaka E, Rapti K, Mavroidis M, Tsangaris G, Maris A, Weisleder N, Capetanaki Y (2005) Alterations in the heart mitochondrial proteome in a desmin null heart failure model. J Mol Cell Cardiol 38(3):461–474. doi:10.1016/j.yjmcc.2004.12.008

Francis F, Roy S, Brady ST, Black MM (2005) Transport of neurofilaments in growing axons requires microtubules but not actin filaments. J Neurosci Res 79(4):442–450. doi:10.1002/jnr.20399

Fu DJ, Thomson C, Lunny DP, Dopping-Hepenstal PJ, McGrath JA, Smith FJ, McLean WH, Pedrioli DM (2014) Keratin 9 is required for the structural integrity and terminal differentiation of the palmoplantar epidermis. J Invest Dermatol 134(3):754–763. doi:10.1038/jid.2013.356

Fuchs E, Green H (1980) Changes in keratin gene expression during terminal differentiation of the keratinocyte. Cell 19(4):1033–1042

Ganay T, Boizot A, Burrer R, Chauvin JP, Bomont P (2011) Sensory-motor deficits and neurofilament disorganization in gigaxonin-null mice. Mol Neurodegener 6:25. doi:10.1186/1750-1326-6-25

Garcia ML, Lobsiger CS, Shah SB, Deerinck TJ, Crum J, Young D, Ward CM, TO C, Gotow T, Uchiyama Y, Ellisman MH, Calcutt NA, Cleveland DW (2003) NF-M is an essential target for the myelin-directed "outside-in" signaling cascade that mediates radial axonal growth. J Cell Biol 163(5):1011–1020. doi:10.1083/jcb.200308159

Garcia ML, Rao MV, Fujimoto J, Garcia VB, Shah SB, Crum J, Gotow T, Uchiyama Y, Ellisman M, Calcutt NA, Cleveland DW (2009) Phosphorylation of highly conserved neurofilament medium KSP repeats is not required for myelin-dependent radial axonal growth. J Neurosci 29(5):1277–1284. doi:10.1523/jneurosci.3765-08.2009

Garcia-Pelagio KP, Muriel J, O'Neill A, Desmond PF, Lovering RM, Lund L, Bond M, Bloch RJ (2015) Myopathic changes in murine skeletal muscle lacking synemin. Am J Physiol Cell Physiol 308(6):C448–C462. doi:10.1152/ajpcell.00331.2014

Gilbert S, Loranger A, Marceau N (2004) Keratins modulate c-Flip/extracellular signal-regulated kinase 1 and 2 antiapoptotic signaling in simple epithelial cells. Mol Cell Biol 24(16):7072–7081. doi:10.1128/MCB.24.16.7072-7081.2004

Goldman RD, Khuon S, Chou YH, Opal P, Steinert PM (1996) The function of intermediate filaments in cell shape and cytoskeletal integrity. J Cell Biol 134(4):971–983

Goldstein J, Horsley V (2012) Home sweet home: skin stem cell niches. Cell Mol Life Sci 69(15):2573–2582. doi:10.1007/s00018-012-0943-3

Gomi H, Yokoyama T, Fujimoto K, Ikeda T, Katoh A, Itoh T, Itohara S (1995) Mice devoid of the glial fibrillary acidic protein develop normally and are susceptible to scrapie prions. Neuron 14(1):29–41

Goodall MH, Ward CW, Pratt SJ, Bloch RJ, Lovering RM (2012) Structural and functional evaluation of branched myofibers lacking intermediate filaments. Am J Physiol Cell Physiol 303(2):C224–C232. doi:10.1152/ajpcell.00136.2012

Goto H, Yasui Y, Kawajiri A, Nigg EA, Terada Y, Tatsuka M, Nagata K, Inagaki M (2003) Aurora-B regulates the cleavage furrow-specific vimentin phosphorylation in the cytokinetic process. J Biol Chem 278(10):8526–8530. doi:10.1074/jbc.M210892200

Grimm SL, Bu W, Longley MA, Roop DR, Li Y, Rosen JM (2006) Keratin 6 is not essential for mammary gland development. Breast Cancer Res 8(3):R29. doi:10.1186/bcr1504

Gruenbaum Y, Foisner R (2015) Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. Annu Rev Biochem 84:131–164. doi:10.1146/annurev-biochem-060614-034115

Guldiken N, Zhou Q, Kucukoglu O, Rehm M, Levada K, Gross A, Kwan R, James LP, Trautwein C, Omary MB, Strnad P (2015) Human keratin 8 variants promote mouse acetaminophen hepatotoxicity coupled with c-jun amino-terminal kinase activation and protein adduct formation. Hepatology 62(3):876–886. doi:10.1002/hep.27891

Hagemann TL, Connor JX, Messing A (2006) Alexander disease-associated glial fibrillary acidic protein mutations in mice induce Rosenthal fiber formation and a white matter stress response. J Neurosci 26(43):11162–11173. doi:10.1523/JNEUROSCI.3260-06.2006

Hagemann TL, Boelens WC, Wawrousek EF, Messing A (2009) Suppression of GFAP toxicity by alphaB-crystallin in mouse models of Alexander disease. Hum Mol Genet 18(7):1190–1199. doi:10.1093/hmg/ddp013

Hagemann TL, Paylor R, Messing A (2013) Deficits in adult neurogenesis, contextual fear conditioning, and spatial learning in a Gfap mutant mouse model of Alexander disease. J Neurosci 33(47):18698–18706. doi:10.1523/jneurosci.3693-13.2013

Hanbury R, Ling ZD, Wuu J, Kordower JH (2003) GFAP knockout mice have increased levels of GDNF that protect striatal neurons from metabolic and excitotoxic insults. J Comp Neurol 461(3):307–316. doi:10.1002/cne.10667

Hapiak V, Hresko MC, Schriefer LA, Saiyasisongkhram K, Bercher M, Plenefisch J (2003) mua-6, a gene required for tissue integrity in *Caenorhabditis elegans*, encodes a cytoplasmic intermediate filament. Dev Biol 263(2):330–342

Hardie DG (2007) AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. Nat Rev Mol Cell Biol 8(10):774–785. doi:10.1038/nrm2249

Helenius TO, Misiorek JO, Nystrom JH, Fortelius LE, Habtezion A, Liao J, Asghar MN, Zhang H, Azhar S, Omary MB, Toivola DM (2015) Keratin 8 absence down-regulates colonocyte HMGCS2 and modulates colonic ketogenesis and energy metabolism. Mol Biol Cell 26(12):2298–2310. doi:10.1091/mbc.E14-02-0736

Hering L, Bouameur JE, Reichelt J, Magin TM, Mayer G (2016) Novel origin of lamin-derived cytoplasmic intermediate filaments in tardigrades. Elife 5:e11117. doi:10.7554/eLife.11117

Herrmann H, Aebi U (2004) Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular Scaffolds. Annu Rev Biochem 73:749–789. doi:10.1146/annurev.biochem.73.011303.073823

Herrmann H, Bar H, Kreplak L, Strelkov SV, Aebi U (2007) Intermediate filaments: from cell architecture to nanomechanics. Nat Rev Mol Cell Biol 8(7):562–573. doi:10.1038/nrm2197

Herrmann H, Strelkov SV, Burkhard P, Aebi U (2009) Intermediate filaments: primary determinants of cell architecture and plasticity. J Clin Invest 119(7):1772–1783. doi:10.1172/JCI38214

Hesse M, Franz T, Tamai Y, Taketo MM, Magin TM (2000) Targeted deletion of keratins 18 and 19 leads to trophoblast fragility and early embryonic lethality. EMBO J 19(19):5060–5070. doi:10.1093/emboj/19.19.5060

Hesse M, Grund C, Herrmann H, Brohl D, Franz T, Omary MB, Magin TM (2007) A mutation of keratin 18 within the coil 1A consensus motif causes widespread keratin aggregation but cell type-restricted lethality in mice. Exp Cell Res 313(14):3127–3140. doi:10.1016/j.yexcr.2007.05.019

Hobbs RP, DePianto DJ, Jacob JT, Han MC, Chung BM, Batazzi AS, Poll BG, Guo Y, Han J, Ong S, Zheng W, Taube JM, Cihakova D, Wan F, Coulombe PA (2015) Keratin-dependent regulation of Aire and gene expression in skin tumor keratinocytes. Nat Genet 47(8):933–938. doi:10.1038/ng.3355

Hol EM, Pekny M (2015) Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system. Curr Opin Cell Biol 32:121–130. doi:10.1016/j.ceb.2015.02.004

Irvine AD, Corden LD, Swensson O, Swensson B, Moore JE, Frazer DG, Smith FJ, Knowlton RG, Christophers E, Rochels R, Uitto J, McLean WH (1997) Mutations in cornea-specific keratin K3 or K12 genes cause Meesmann's corneal dystrophy. Nat Genet 16(2):184–187. doi:10.1038/ng0697-184

Ivaska J, Pallari HM, Nevo J, Eriksson JE (2007) Novel functions of vimentin in cell adhesion, migration, and signaling. Exp Cell Res 313(10):2050–2062. doi:10.1016/j.yexcr.2007.03.040

Izmiryan A, Cheraud Y, Khanamiryan L, Leterrier JF, Federici T, Peltekian E, Moura-Neto V, Paulin D, Li Z, Xue ZG (2006) Different expression of synemin isoforms in glia and neurons during nervous system development. Glia 54(3):204–213. doi:10.1002/glia.20378

Izmiryan A, Franco CA, Paulin D, Li Z, Xue Z (2009) Synemin isoforms during mouse development: multiplicity of partners in vascular and neuronal systems. Exp Cell Res 315(5):769–783. doi:10.1016/j.yexcr.2008.12.009

Jacomy H, Zhu Q, Couillard-Despres S, Beaulieu JM, Julien JP (1999) Disruption of type IV intermediate filament network in mice lacking the neurofilament medium and heavy subunits. J Neurochem 73(3):972–984

Jaeschke H, McGill MR, Ramachandran A (2012) Oxidant stress, mitochondria, and cell death mechanisms in drug-induced liver injury: lessons learned from acetaminophen hepatotoxicity. Drug Metab Rev 44(1):88–106. doi:10.3109/03602532.2011.602688

Jahnel O, Hoffmann B, Merkel R, Bossinger O, Leube RE (2016) Mechanical probing of the intermediate filament-rich *Caenorhabditis elegans* intestine. Methods Enzymol 568:681–706. doi:10.1016/bs.mie.2015.08.030

Jany PL, Hagemann TL, Messing A (2013) GFAP expression as an indicator of disease severity in mouse models of Alexander disease. ASN Neuro 5(1):e00109. doi:10.1042/an20130003

Jaquemar D, Kupriyanov S, Wankell M, Avis J, Benirschke K, Baribault H, Oshima RG (2003) Keratin 8 protection of placental barrier function. J Cell Biol 161(4):749–756. doi:10.1083/jcb.200210004

Jing R, Wilhelmsson U, Goodwill W, Li L, Pan Y, Pekny M, Skalli O (2007) Synemin is expressed in reactive astrocytes in neurotrauma and interacts differentially with vimentin and GFAP intermediate filament networks. J Cell Sci 120(Pt 7):1267–1277. doi:10.1242/jcs.03423

Kamphuis W, Kooijman L, Orre M, Stassen O, Pekny M, Hol EM (2015) GFAP and vimentin deficiency alters gene expression in astrocytes and microglia in wild-type mice and changes the transcriptional response of reactive glia in mouse model for Alzheimer's disease. Glia 63(6):1036–1056. doi:10.1002/glia.22800

Kao WW, Liu CY, Converse RL, Shiraishi A, Kao CW, Ishizaki M, Doetschman T, Duffy J (1996) Keratin 12-deficient mice have fragile corneal epithelia. Invest Ophthalmol Vis Sci 37(13):2572–2584

Karabinos A, Schmidt H, Harborth J, Schnabel R, Weber K (2001) Essential roles for four cytoplasmic intermediate filament proteins in *Caenorhabditis elegans* development. Proc Natl Acad Sci U S A 98(14):7863–7868. doi:10.1073/pnas.121169998

Karabinos A, Schulze E, Schunemann J, Parry DAD, Weber K (2003) In vivo and in vitro evidence that the four essential intermediate filament (IF) proteins A1, A2, A3 and B1 of the nematode *Caenorhabditis elegans* form an obligate heteropolymeric IF system. J Mol Biol 333(2):307–319

Karabinos A, Schunemann J, Weber K (2004) Most genes encoding cytoplasmic intermediate fila-ment (IF) proteins of the nematode *Caenorhabditis elegans* are required in late embryogenesis. Eur J Cell Biol 83(9):457–468. doi:10.1078/0171-9335-00407

Kawajiri A, Yasui Y, Goto H, Tatsuka M, Takahashi M, Nagata K, Inagaki M (2003) Functional significance of the specific sites phosphorylated in desmin at cleavage furrow: Aurora-B may phosphorylate and regulate type III intermediate filaments during cytokinesis coordinatedly with Rho-kinase. Mol Biol Cell 14(4):1489–1500. doi:10.1091/mbc.E02-09-0612

Kemp MW, Edwards B, Burgess M, Clarke WT, Nicholson G, Parry DAD, Davies KE (2009) Syncoilin isoform organization and differential expression in murine striated muscle. J Struct Biol 165(3):196–203. doi:10.1016/j.jsb.2008.11.002

Kerns ML, DePianto D, Dinkova-Kostova AT, Talalay P, Coulombe PA (2007) Reprogramming of keratin biosynthesis by sulforaphane restores skin integrity in epidermolysis bullosa simplex. Proc Natl Acad Sci U S A 104(36):14460–14465. doi:10.1073/pnas.0706486104

Kerns M, DePianto D, Yamamoto M, Coulombe PA (2010) Differential modulation of keratin expression by sulforaphane occurs via Nrf2-dependent and -independent pathways in skin epi-thelia. Mol Biol Cell 21(23):4068–4075. doi:10.1091/mbc.E10-02-0153

Kim S, Wong P, Coulombe PA (2006) A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. Nature 441(7091):362–365. doi:10.1038/nature04659

Kolotuev I, Hyenne V, Schwab Y, Rodriguez D, Labouesse M (2013) A pathway for unicellular tube extension depending on the lymphatic vessel determinant Prox1 and on osmoregulation. Nat Cell Biol 15(2):157–168. doi:10.1038/ncb2662

Kong J, Tung VW, Aghajanian J, Xu Z (1998) Antagonistic roles of neurofilament subunits NF-H and NF-M against NF-L in shaping dendritic arborization in spinal motor neurons. J Cell Biol 140(5):1167–1176

Kriz J, Zhu Q, Julien JP, Padjen AL (2000) Electrophysiological properties of axons in mice lack-ing neurofilament subunit genes: disparity between conduction velocity and axon diameter in absence of NF-H. Brain Res 885(1):32–44

Kroger C, Vijayaraj P, Reuter U, Windoffer R, Simmons D, Heukamp L, Leube R, Magin TM (2011) Placental vasculogenesis is regulated by keratin-mediated hyperoxia in murine decidual tissues. Am J Pathol 178(4):1578–1590. doi:10.1016/j.ajpath.2010.12.055

Kroger C, Loschke F, Schwarz N, Windoffer R, Leube RE, Magin TM (2013) Keratins control intercellular adhesion involving PKC-alpha-mediated desmoplakin phosphorylation. J Cell Biol 201(5):681–692. doi:10.1083/jcb.201208162

Ku NO, Omary MB (1995) Identification and mutational analysis of the glycosylation sites of human keratin 18. J Biol Chem 270(20):11820–11827

Ku NO, Omary MB (2006) A disease- and phosphorylation-related nonmechanical function for keratin 8. J Cell Biol 174(1):115–125. doi:10.1083/jcb.200602146

Ku NO, Michie S, Resurreccion EZ, Broome RL, Omary MB (2002) Keratin binding to 14-3-3 proteins modulates keratin filaments and hepatocyte mitotic progression. Proc Natl Acad Sci U S A 99(7):4373–4378. doi:10.1073/pnas.072624299

Ku NO, Fu H, Omary MB (2004) Raf-1 activation disrupts its binding to keratins during cell stress. J Cell Biol 166(4):479–485. doi:10.1083/jcb.200402051

Ku NO, Strnad P, Zhong BH, Tao GZ, Omary MB (2007) Keratins let liver live: mutations predis-pose to liver disease and crosslinking generates Mallory-Denk bodies. Hepatology 46(5):1639–1649. doi:10.1002/hep.21976

Ku NO, Toivola DM, Strnad P, Omary MB (2010) Cytoskeletal keratin glycosylation protects epithelial tissue from injury. Nat Cell Biol 12(9):876–885. doi:10.1038/ncb2091

Kucukoglu O, Guldiken N, Chen Y, Usachov V, El-Heliebi A, Haybaeck J, Denk H, Trautwein C, Strnad P (2014) High-fat diet triggers Mallory-Denk body formation through misfolding and crosslinking of excess keratin 8. Hepatology 60(1):169–178. doi:10.1002/hep.27068

Kuga T, Kume H, Kawasaki N, Sato M, Adachi J, Shiromizu T, Hoshino I, Nishimori T, Matsubara H, Tomonaga T (2013) A novel mechanism of keratin cytoskeleton organization through casein kinase Ialpha and FAM83H in colorectal cancer. J Cell Sci 126(Pt 20):4721–4731. doi:10.1242/jcs.129684

Kumar V, Pedroza LA, Mace EM, Seeholzer S, Cotsarelis G, Condino-Neto A, Payne AS, Orange JS (2011) The autoimmune regulator (AIRE), which is defective in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients, is expressed in human epidermal and follicular keratinocytes and associates with the intermediate filament protein cytokeratin 17. Am J Pathol 178(3):983–988. doi:10.1016/j.ajpath.2010.12.007

Kumar V, Bouameur JE, Bar J, Rice RH, Hornig-Do HT, Roop DR, Schwarz N, Brodesser S, Thiering S, Leube RE, Wiesner RJ, Brazel CB, Heller S, Binder H, Loffler-Wirth H, Seibel P, Magin TM (2015) A keratin scaffold regulates epidermal barrier formation, mitochondrial lipid composition, and activity. J Cell Biol 211(5):1057–1075. doi:10.1083/jcb.201404147

Kumari SS, Varadaraj K (2014) Aquaporin 0 plays a pivotal role in refractive index gradient development in mammalian eye lens to prevent spherical aberration. Biochem Biophys Res Commun 452(4):986–991. doi:10.1016/j.bbrc.2014.09.032

Kwan R, Chen L, Looi K, Tao GZ, Weerasinghe SV, Snider NT, Conti MA, Adelstein RS, Xie Q, Omary MB (2015) PKC412 normalizes mutation-related keratin filament disruption and hepatic injury in mice by promoting keratin-myosin binding. Hepatology. doi:10.1002/hep.27965

Lariviere RC, Julien JP (2004) Functions of intermediate filaments in neuronal development and disease. J Neurobiol 58(1):131–148. doi:10.1002/neu.10270

Larouche D, Tong X, Fradette J, Coulombe PA, Germain L (2008) Vibrissa hair bulge houses two populations of skin epithelial stem cells distinct by their keratin profile. FASEB J 22(5):1404–1415. doi:10.1096/fj.07-8109com

Larsson A, Wilhelmsson U, Pekna M, Pekny M (2004) Increased cell proliferation and neurogenesis in the hippocampal dentate gyrus of old GFAP(−/−)Vim(−/−) mice. Neurochem Res 29(11):2069–2073

Le Henaff C, Da Cunha MF, Hatton A, Tondelier D, Marty C, Collet C, Zarka M, Geoffroy V, Zatloukal K, Laplantine E, Edelman A, Sermet-Gaudelus I, Marie PJ (2016) Genetic deletion of keratin 8 corrects the altered bone formation and osteopenia in a mouse model of cystic fibrosis. Hum Mol Genet. doi:10.1093/hmg/ddw009

Lebkuechner I, Wilhelmsson U, Mollerstrom E, Pekna M, Pekny M (2015) Heterogeneity of Notch signaling in astrocytes and the effects of GFAP and vimentin deficiency. J Neurochem 135(2):234–248. doi:10.1111/jnc.13213

Lee MK, Xu Z, Wong PC, Cleveland DW (1993) Neurofilaments are obligate heteropolymers in vivo. J Cell Biol 122(6):1337–1350

Lee J, Jang KH, Kim H, Lim Y, Kim S, Yoon HN, Chung IK, Roth J, Ku NO (2013) Predisposition to apoptosis in keratin 8-null liver is related to inactivation of NF-kappaB and SAPKs but not decreased c-Flip. Biol Open 2(7):695–702. doi:10.1242/bio.20134606

Lepinoux-Chambaud C, Eyer J (2013) Review on intermediate filaments of the nervous system and their pathological alterations. Histochem Cell Biol 140(1):13–22. doi:10.1007/s00418-013-1101-1

Lessard JC, Coulombe PA (2012) Keratin 16-null mice develop palmoplantar keratoderma, a hallmark feature of pachyonychia congenita and related disorders. J Invest Dermatol 132(5):1384–1391. doi:10.1038/jid.2012.6

Lessard JC, Pina-Paz S, Rotty JD, Hickerson RP, Kaspar RL, Balmain A, Coulombe PA (2013) Keratin 16 regulates innate immunity in response to epidermal barrier breach. Proc Natl Acad Sci U S A 110(48):19537–19542. doi:10.1073/pnas.1309576110

Li Z, Colucci-Guyon E, Pincon-Raymond M, Mericskay M, Pournin S, Paulin D, Babinet C (1996) Cardiovascular lesions and skeletal myopathy in mice lacking desmin. Dev Biol 175(2):362–366. doi:10.1006/dbio.1996.0122

Li Z, Mericskay M, Agbulut O, Butler-Browne G, Carlsson L, Thornell LE, Babinet C, Paulin D (1997) Desmin is essential for the tensile strength and integrity of myofibrils but not for myogenic commitment, differentiation, and fusion of skeletal muscle. J Cell Biol 139(1):129–144

Li L, Lundkvist A, Andersson D, Wilhelmsson U, Nagai N, Pardo AC, Nodin C, Stahlberg A, Aprico K, Larsson K, Yabe T, Moons L, Fotheringham A, Davies I, Carmeliet P, Schwartz JP, Pekna M, Kubista M, Blomstrand F, Maragakis N, Nilsson M, Pekny M (2008) Protective role of reactive astrocytes in brain ischemia. J Cereb Blood Flow Metab 28(3):468–481. doi:10.1038/sj.jcbfm.9600546

Li Z, Parlakian A, Coletti D, Alonso-Martin S, Hourde C, Joanne P, Gao-Li J, Blanc J, Ferry A, Paulin D, Xue Z, Agbulut O (2014) Synemin acts as a regulator of signalling molecules during skeletal muscle hypertrophy. J Cell Sci 127(Pt 21):4589–4601. doi:10.1242/jcs.143164

Lieber RL, Schmitz MC, Mishra DK, Friden J (1994) Contractile and cellular remodeling in rabbit skeletal muscle after cyclic eccentric contractions. J Appl Physiol (1985) 77(4):1926–1934

Liedtke W, Edelmann W, Bieri PL, Chiu FC, Cowan NJ, Kucherlapati R, Raine CS (1996) GFAP is necessary for the integrity of CNS white matter architecture and long-term maintenance of myelination. Neuron 17(4):607–615

Liem RK, Messing A (2009) Dysfunctions of neuronal and glial intermediate filaments in disease. J Clin Invest 119(7):1814–1824. doi:10.1172/jci38003

Lindqvist J, Wistbacka N, Eriksson JE (2016) Studying nestin and its interrelationship with Cdk5. Methods Enzymol 568:509–535. doi:10.1016/bs.mie.2015.09.019

Liu Z, Li Y, Cui Y, Roberts C, Lu M, Wilhelmsson U, Pekny M, Chopp M (2014) Beneficial effects of gfap/vimentin reactive astrocytes for axonal remodeling and motor behavioral recovery in mice after stroke. Glia 62(12):2022–2033. doi:10.1002/glia.22723

Lloyd C, Yu QC, Cheng J, Turksen K, Degenstein L, Hutton E, Fuchs E (1995) The basal keratin network of stratified squamous epithelia: defining K15 function in the absence of K14. J Cell Biol 129(5):1329–1344

Loranger A, Duclos S, Grenier A, Price J, Wilson-Heiner M, Baribault H, Marceau N (1997) Simple epithelium keratins are required for maintenance of hepatocyte integrity. Am J Pathol 151(6):1673–1683

Loschke F, Seltmann K, Bouameur JE, Magin TM (2015) Regulation of keratin network organization. Curr Opin Cell Biol 32:56–64. doi:10.1016/j.ceb.2014.12.006

Loschke F, Homberg M, Magin TM (2016) Keratin isotypes control desmosome stability and dynamics through PKCalpha. J Invest Dermatol 136(1):202–213. doi:10.1038/JID.2015.403

Lovering RM, O'Neill A, Muriel JM, Prosser BL, Strong J, Bloch RJ (2011) Physiology, structure, and susceptibility to injury of skeletal muscle in mice lacking keratin 19-based and desmin-based intermediate filaments. Am J Physiol Cell Physiol 300(4):C803–C813. doi:10.1152/ajpcell.00394.2010

Lu H, Zimek A, Chen J, Hesse M, Bussow H, Weber K, Magin TM (2006) Keratin 5 knockout mice reveal plasticity of keratin expression in the corneal epithelium. Eur J Cell Biol 85(8):803–811. doi:10.1016/j.ejcb.2006.04.001

Lyle S, Christofidou-Solomidou M, Liu Y, Elder DE, Albelda S, Cotsarelis G (1998) The C8/144B monoclonal antibody recognizes cytokeratin 15 and defines the location of human hair follicle stem cells. J Cell Sci 111(Pt 21):3179–3188

Magin TM, Schroder R, Leitgeb S, Wanninger F, Zatloukal K, Grund C, Melton DW (1998) Lessons from keratin 18 knockout mice: formation of novel keratin filaments, secondary loss of keratin 7 and accumulation of liver-specific keratin 8-positive aggregates. J Cell Biol 140(6):1441–1451

Magin TM, Hesse M, Meier-Bornheim R, Reichelt J (2004a) Developing mouse models to study intermediate filament function. Methods Cell Biol 78:65–94

Magin TM, Reichelt J, Hatzfeld M (2004b) Emerging functions: diseases and animal models reshape our view of the cytoskeleton. Exp Cell Res 301(1):91–102. doi:10.1016/j.yexcr.2004.08.018

Mahammad S, Murthy SN, Didonna A, Grin B, Israeli E, Perrot R, Bomont P, Julien JP, Kuczmarski E, Opal P, Goldman RD (2013) Giant axonal neuropathy-associated gigaxonin mutations impair intermediate filament protein degradation. J Clin Invest 123(5):1964–1975. doi:10.1172/JCI66387

Maison S, Liberman LD, Liberman MC (2016) Type II cochlear ganglion neurons do not drive the olivocochlear reflex: re-examination of the cochlear phenotype in peripherin knock-out mice. eNeuro 3(4). pii: ENEURO.0207-16.2016. doi:10.1523/ENEURO.0207-16.2016

Mann DL (2003) Stress-activated cytokines and the heart: from adaptation to maladaptation. Annu Rev Physiol 65:81–101. doi:10.1146/annurev.physiol.65.092101.142249

Marszalek JR, Williamson TL, Lee MK, Xu Z, Hoffman PN, Becher MW, TO C, Cleveland DW (1996) Neurofilament subunit NF-H modulates axonal diameter by selectively slowing neurofilament transport. J Cell Biol 135(3):711–724

Mathew J, Loranger A, Gilbert S, Faure R, Marceau N (2013) Keratin 8/18 regulation of glucose metabolism in normal versus cancerous hepatic cells through differential modulation of hexokinase status and insulin signaling. Exp Cell Res 319(4):474–486. doi:10.1016/j.yexcr.2012.11.011

Matsui T, Amagai M (2015) Dissecting the formation, structure and barrier function of the stratum corneum. Int Immunol 27(6):269–280. doi:10.1093/intimm/dxv013

Matsuyama M, Tanaka H, Inoko A, Goto H, Yonemura S, Kobori K, Hayashi Y, Kondo E, Itohara S, Izawa I, Inagaki M (2013) Defect of mitotic vimentin phosphorylation causes microophthalmia and cataract via aneuploidy and senescence in lens epithelial cells. J Biol Chem 288(50):35626–35635. doi:10.1074/jbc.M113.514737

Mavroidis M, Capetanaki Y (2002) Extensive induction of important mediators of fibrosis and dystrophic calcification in desmin-deficient cardiomyopathy. Am J Pathol 160(3):943–952. doi:10.1016/s0002-9440(10)64916-4

Mavroidis M, Panagopoulou P, Kostavasili I, Weisleder N, Capetanaki Y (2008) A missense mutation in desmin tail domain linked to human dilated cardiomyopathy promotes cleavage of the head domain and abolishes its Z-disc localization. FASEB J 22(9):3318–3327. doi:10.1096/fj.07-088724

Mavroidis M, Davos CH, Psarras S, Varela ANCA, Katsimpoulas M, Kostavasili I, Maasch C, Vater A, van Tintelen JP, Capetanaki Y (2015) Complement system modulation as a target for treatment of arrhythmogenic cardiomyopathy. Basic Res Cardiol 110(3):27. doi:10.1007/s00395-015-0485-6

McAleer MA, Pohler E, Smith FJ, Wilson NJ, Cole C, MacGowan S, Koetsier JL, Godsel LM, Harmon RM, Gruber R, Crumrine D, Elias PM, McDermott M, Butler K, Broderick A, Sarig O, Sprecher E, Green KJ, McLean WH, Irvine AD (2015) Severe dermatitis, multiple allergies, and metabolic wasting syndrome caused by a novel mutation in the N-terminal plakin domain of desmoplakin. J Allergy Clin Immunol 136(5):1268–1276. doi:10.1016/j.jaci.2015.05.002

McCall MA, Gregg RG, Behringer RR, Brenner M, Delaney CL, Galbreath EJ, Zhang CL, Pearce RA, Chiu SY, Messing A (1996) Targeted deletion in astrocyte intermediate filament (Gfap) alters neuronal physiology. Proc Natl Acad Sci U S A 93(13):6361–6366

McCullagh KJ, Edwards B, Kemp MW, Giles LC, Burgess M, Davies KE (2008) Analysis of skeletal muscle function in the C57BL6/SV129 syncoilin knockout mouse. Mamm Genome 19(5):339–351. doi:10.1007/s00335-008-9120-2

McGowan KM, Tong X, Colucci-Guyon E, Langa F, Babinet C, Coulombe PA (2002) Keratin 17 null mice exhibit age- and strain-dependent alopecia. Genes Dev 16(11):1412–1422. doi:10.1101/gad.979502

McLean WH, Hansen CD, Eliason MJ, Smith FJ (2011) The phenotypic and molecular genetic features of pachyonychia congenita. J Invest Dermatol 131(5):1015–1017. doi:10.1038/jid.2011.59

Meier J, Couillard-Despres S, Jacomy H, Gravel C, Julien JP (1999) Extra neurofilament NF-L subunits rescue motor neuron disease caused by overexpression of the human NF-H gene in mice. J Neuropathol Exp Neurol 58(10):1099–1110

Messing A, Head MW, Galles K, Galbreath EJ, Goldman JE, Brenner M (1998) Fatal encephalopathy with astrocyte inclusions in GFAP transgenic mice. Am J Pathol 152(2):391–398

Millecamps S, Gowing G, Corti O, Mallet J, Julien JP (2007) Conditional NF-L transgene expression in mice for in vivo analysis of turnover and transport rate of neurofilaments. J Neurosci 27(18):4947–4956. doi:10.1523/jneurosci.5299-06.2007

Milner DJ, Weitzer G, Tran D, Bradley A, Capetanaki Y (1996) Disruption of muscle architecture and myocardial degeneration in mice lacking desmin. J Cell Biol 134(5):1255–1270

Milner DJ, Taffet GE, Wang X, Pham T, Tamura T, Hartley C, Gerdes AM, Capetanaki Y (1999) The absence of desmin leads to cardiomyocyte hypertrophy and cardiac dilation with compromised systolic function. J Mol Cell Cardiol 31(11):2063–2076

Mizuno Y, Guyon JR, Okamoto K, Kunkel LM (2009) Expression of synemin in the mouse spinal cord. Muscle Nerve 39(5):634–641. doi:10.1002/mus.21221

Mohseni P, Sung HK, Murphy AJ, Laliberte CL, Pallari HM, Henkelman M, Georgiou J, Xie G, Quaggin SE, Thorner PS, Eriksson JE, Nagy A (2011) Nestin is not essential for development of the CNS but required for dispersion of acetylcholine receptor clusters at the area of neuromuscular junctions. J Neurosci 31(32):11547–11552. doi:10.1523/JNEUROSCI.4396-10.2011

Moll I, Kuhn C, Moll R (1995) Cytokeratin 20 is a general marker of cutaneous Merkel cells while certain neuronal proteins are absent. J Invest Dermatol 104(6):910–915

Moll R, Divo M, Langbein L (2008) The human keratins: biology and pathology. Histochem Cell Biol 129(6):705–733. doi:10.1007/s00418-008-0435-6

Mor-Vaknin N, Legendre M, Yu Y, Serezani CH, Garg SK, Jatzek A, Swanson MD, Gonzalez-Hernandez MJ, Teitz-Tennenbaum S, Punturieri A, Engleberg NC, Banerjee R, Peters-Golden M, Kao JY, Markovitz DM (2013) Murine colitis is mediated by vimentin. Sci Rep 3:1045. doi:10.1038/srep01045

Muller M, Bhattacharya SS, Moore T, Prescott Q, Wedig T, Herrmann H, Magin TM (2009) Dominant cataract formation in association with a vimentin assembly disrupting mutation. Hum Mol Genet 18(6):1052–1057. doi:10.1093/hmg/ddn440

Nakagawa T, Zhu H, Morishima N, Li E, Xu J, Yankner BA, Yuan J (2000) Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid-beta. Nature 403(6765):98–103. doi:10.1038/47513

Nakamichi I, Toivola DM, Strnad P, Michie SA, Oshima RG, Baribault H, Omary MB (2005) Keratin 8 overexpression promotes mouse Mallory body formation. J Cell Biol 171(6):931–937. doi:10.1083/jcb.200507093

Nestle FO, Di Meglio P, Qin JZ, Nickoloff BJ (2009) Skin immune sentinels in health and disease. Nat Rev Immunol 9(10):679–691. doi:10.1038/nri2622

Nguyen MD, Lariviere RC, Julien JP (2001) Deregulation of Cdk5 in a mouse model of ALS: toxicity alleviated by perikaryal neurofilament inclusions. Neuron 30(1):135–147

Nguyen MD, Shu T, Sanada K, Lariviere RC, Tseng HC, Park SK, Julien JP, Tsai LH (2004) A NUDEL-dependent mechanism of neurofilament assembly regulates the integrity of CNS neurons. Nat Cell Biol 6(7):595–608. doi:10.1038/ncb1139

Nieminen M, Henttinen T, Merinen M, Marttila-Ichihara F, Eriksson JE, Jalkanen S (2006) Vimentin function in lymphocyte adhesion and transcellular migration. Nat Cell Biol 8(2):156–162. doi:10.1038/ncb1355

O'Neill A, Williams MW, Resneck WG, Milner DJ, Capetanaki Y, Bloch RJ (2002) Sarcolemmal organization in skeletal muscle lacking desmin: evidence for cytokeratins associated with the membrane skeleton at costameres. Mol Biol Cell 13(7):2347–2359. doi:10.1091/mbc.01-12-0576

Odaka C, Loranger A, Takizawa K, Ouellet M, Tremblay MJ, Murata S, Inoko A, Inagaki M, Marceau N (2013) Keratin 8 is required for the maintenance of architectural structure in thymus epithelium. PLoS One 8(9):e75101. doi:10.1371/journal.pone.0075101

Ohara O, Gahara Y, Miyake T, Teraoka H, Kitamura T (1993) Neurofilament deficiency in quail caused by nonsense mutation in neurofilament-L gene. J Cell Biol 121(2):387–395

Oji V, Eckl KM, Aufenvenne K, Natebus M, Tarinski T, Ackermann K, Seller N, Metze D, Nurnberg G, Folster-Holst R, Schafer-Korting M, Hausser I, Traupe H, Hennies HC (2010) Loss of corneodesmosin leads to severe skin barrier defect, pruritus, and atopy: unraveling the peeling skin disease. Am J Hum Genet 87(2):274–281. doi:10.1016/j.ajhg.2010.07.005

Omary MB, Ku NO, Strnad P, Hanada S (2009) Toward unraveling the complexity of simple epithelial keratins in human disease. J Clin Invest 119(7):1794–1805. doi:10.1172/JCI37762

Ouellet T, Lussier M, Babai F, Lapointe L, Royal A (1990) Differential expression of the epidermal K1 and K10 keratin genes during mouse embryo development. Biochem Cell Biol 68(2):448–453

Paladini RD, Coulombe PA (1998) Directed expression of keratin 16 to the progenitor basal cells of transgenic mouse skin delays skin maturation. J Cell Biol 142(4):1035–1051

Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, Goudie DR, Sandilands A, Campbell LE, Smith FJ, O'Regan GM, Watson RM, Cecil JE, Bale SJ, Compton JG, DiGiovanna JJ, Fleckman P, Lewis-Jones S, Arseculeratne G, Sergeant A, Munro CS, El Houate B, McElreavey K, Halkjaer LB, Bisgaard H, Mukhopadhyay S, McLean WH (2006) Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat Genet 38(4):441–446. doi:10.1038/ng1767

Panagopoulou P, Davos CH, Milner DJ, Varela E, Cameron J, Mann DL, Capetanaki Y (2008) Desmin mediates TNF-alpha-induced aggregate formation and intercalated disk reorganization in heart failure. J Cell Biol 181(5):761–775. doi:10.1083/jcb.200710049

Papathanasiou S, Rickelt S, Soriano ME, Schips TG, Maier HJ, Davos CH, Varela A, Kaklamanis L, Mann DL, Capetanaki Y (2015) Tumor necrosis factor-alpha confers cardioprotection through ectopic expression of keratins K8 and K18. Nat Med 21(9):1076–1084. doi:10.1038/nm.3925

Paramio JM, Segrelles C, Ruiz S, Jorcano JL (2001) Inhibition of protein kinase B (PKB) and PKCzeta mediates keratin K10-induced cell cycle arrest. Mol Cell Biol 21(21):7449–7459. doi:10.1128/MCB.21.21.7449-7459.2001

Park D, Xiang AP, Mao FF, Zhang L, Di CG, Liu XM, Shao Y, Ma BF, Lee JH, Ha KS, Walton N, Lahn BT (2010) Nestin is required for the proper self-renewal of neural stem cells. Stem Cells 28(12):2162–2171. doi:10.1002/stem.541

Pekny M, Nilsson M (2005) Astrocyte activation and reactive gliosis. Glia 50(4):427–434. doi:10.1002/glia.20207

Pekny M, Pekna M (2004) Astrocyte intermediate filaments in CNS pathologies and regeneration. J Pathol 204(4):428–437. doi:10.1002/path.1645

Pekny M, Leveen P, Pekna M, Eliasson C, Berthold CH, Westermark B, Betsholtz C (1995) Mice lacking glial fibrillary acidic protein display astrocytes devoid of intermediate filaments but develop and reproduce normally. EMBO J 14(8):1590–1598

Pekny M, Johansson CB, Eliasson C, Stakeberg J, Wallen A, Perlmann T, Lendahl U, Betsholtz C, Berthold CH, Frisen J (1999) Abnormal reaction to central nervous system injury in mice lacking glial fibrillary acidic protein and vimentin. J Cell Biol 145(3):503–514

Pekny M, Wilhelmsson U, Pekna M (2014) The dual role of astrocyte activation and reactive gliosis. Neurosci Lett 565:30–38. doi:10.1016/j.neulet.2013.12.071

Peter A, Stick R (2015) Evolutionary aspects in intermediate filament proteins. Curr Opin Cell Biol 32:48–55. doi:10.1016/j.ceb.2014.12.009

Peters B, Kirfel J, Bussow H, Vidal M, Magin TM (2001) Complete cytolysis and neonatal lethality in keratin 5 knockout mice reveal its fundamental role in skin integrity and in epidermolysis bullosa simplex. Mol Biol Cell 12(6):1775–1789

Potten CS, Morris RJ (1988) Epithelial stem cells in vivo. J Cell Sci Suppl 10:45–62

Presland RB, Coulombe PA, Eckert RL, Mao-Qiang M, Feingold KR, Elias PM (2004) Barrier function in transgenic mice overexpressing K16, involucrin, and filaggrin in the suprabasal epidermis. J Invest Dermatol 123(3):603–606. doi:10.1111/j.0022-202X.2004.23226.x

Prudner BC, Roy PS, Damron DS, Russell MA (2014) alpha-Synemin localizes to the M-band of the sarcomere through interaction with the M10 region of titin. FEBS Lett 588(24):4625–4630. doi:10.1016/j.febslet.2014.11.001

Ralston E, Lu Z, Biscocho N, Soumaka E, Mavroidis M, Prats C, Lomo T, Capetanaki Y, Ploug T (2006) Blood vessels and desmin control the positioning of nuclei in skeletal muscle fibers. J Cell Physiol 209(3):874–882. doi:10.1002/jcp.20780

Ramms L, Fabris G, Windoffer R, Schwarz N, Springer R, Zhou C, Lazar J, Stiefel S, Hersch N, Schnakenberg U, Magin TM, Leube RE, Merkel R, Hoffmann B (2013) Keratins as the main component for the mechanical integrity of keratinocytes. Proc Natl Acad Sci U S A 110(46):18513–18518. doi:10.1073/pnas.1313491110

Rao MV, Engle LJ, Mohan PS, Yuan A, Qiu D, Cataldo A, Hassinger L, Jacobsen S, Lee VM, Andreadis A, Julien JP, Bridgman PC, Nixon RA (2002a) Myosin Va binding to neurofilaments

is essential for correct myosin Va distribution and transport and neurofilament density. J Cell Biol 159(2):279–290. doi:10.1083/jcb.200205062

Rao MV, Garcia ML, Miyazaki Y, Gotow T, Yuan A, Mattina S, Ward CM, Calcutt NA, Uchiyama Y, Nixon RA, Cleveland DW (2002b) Gene replacement in mice reveals that the heavily phosphorylated tail of neurofilament heavy subunit does not affect axonal caliber or the transit of cargoes in slow axonal transport. J Cell Biol 158(4):681–693. doi:10.1083/jcb.200202037

Rao MV, Campbell J, Yuan A, Kumar A, Gotow T, Uchiyama Y, Nixon RA (2003) The neurofilament middle molecular mass subunit carboxyl-terminal tail domains is essential for the radial growth and cytoskeletal architecture of axons but not for regulating neurofilament transport rate. J Cell Biol 163(5):1021–1031. doi:10.1083/jcb.200308076

Rao MV, Mohan PS, Kumar A, Yuan A, Montagna L, Campbell J, Veeranna EEM, Julien JP, Nixon RA (2011) The myosin Va head domain binds to the neurofilament-L rod and modulates endoplasmic reticulum (ER) content and distribution within axons. PLoS One 6(2):e17087. doi:10.1371/journal.pone.0017087

Reichelt J, Magin TM (2002) Hyperproliferation, induction of c-Myc and 14-3-3sigma, but no cell fragility in keratin-10-null mice. J Cell Sci 115(Pt 13):2639–2650

Reichelt J, Bauer C, Porter R, Lane E, Magin V (1997) Out of balance: consequences of a partial keratin 10 knockout. J Cell Sci 110(Pt 18):2175–2186

Reichelt J, Doering T, Schnetz E, Fartasch M, Sandhoff K, Magin AM (1999) Normal ultrastructure, but altered stratum corneum lipid and protein composition in a mouse model for epidermolytic hyperkeratosis. J Invest Dermatol 113(3):329–334. doi:10.1046/j.1523-1747.1999.00702.x

Reichelt J, Bussow H, Grund C, Magin TM (2001) Formation of a normal epidermis supported by increased stability of keratins 5 and 14 in keratin 10 null mice. Mol Biol Cell 12(6):1557–1568

Reichelt J, Furstenberger G, Magin TM (2004) Loss of keratin 10 leads to mitogen-activated protein kinase (MAPK) activation, increased keratinocyte turnover, and decreased tumor formation in mice. J Invest Dermatol 123(5):973–981. doi:10.1111/j.0022-202X.2004.23426.x

Roth W, Kumar V, Beer HD, Richter M, Wohlenberg C, Reuter U, Thiering S, Staratschek-Jox A, Hofmann A, Kreusch F, Schultze JL, Vogl T, Roth J, Reichelt J, Hausser I, Magin TM (2012) Keratin 1 maintains skin integrity and participates in an inflammatory network in skin through interleukin-18. J Cell Sci 125(Pt 22):5269–5279. doi:10.1242/jcs.116574

Rouleau GA, Clark AW, Rooke K, Pramatarova A, Krizus A, Suchowersky O, Julien JP, Figlewicz D (1996) SOD1 mutation is associated with accumulation of neurofilaments in amyotrophic lateral sclerosis. Ann Neurol 39(1):128–131. doi:10.1002/ana.410390119

Sahlgren CM, Mikhailov A, Hellman J, Chou YH, Lendahl U, Goldman RD, Eriksson JE (2001) Mitotic reorganization of the intermediate filament protein nestin involves phosphorylation by cdc2 kinase. J Biol Chem 276(19):16456–16463. doi:10.1074/jbc.M009669200

Sahlgren CM, Mikhailov A, Vaittinen S, Pallari HM, Kalimo H, Pant HC, Eriksson JE (2003) Cdk5 regulates the organization of nestin and its association with p35. Mol Cell Biol 23(14):5090–5106

Sakaguchi T, Okada M, Kitamura T, Kawasaki K (1993) Reduced diameter and conduction velocity of myelinated fibers in the sciatic nerve of a neurofilament-deficient mutant quail. Neurosci Lett 153(1):65–68

Samuelov L, Sarig O, Harmon RM, Rapaport D, Ishida-Yamamoto A, Isakov O, Koetsier JL, Gat A, Goldberg I, Bergman R, Spiegel R, Eytan O, Geller S, Peleg S, Shomron N, Goh CS, Wilson NJ, Smith FJ, Pohler E, Simpson MA, McLean WH, Irvine AD, Horowitz M, McGrath JA, Green KJ, Sprecher E (2013) Desmoglein 1 deficiency results in severe dermatitis, multiple allergies and metabolic wasting. Nat Genet 45(10):1244–1248. doi:10.1038/ng.2739

Sandilands A, Prescott AR, Wegener A, Zoltoski RK, Hutcheson AM, Masaki S, Kuszak JR, Quinlan RA (2003) Knockout of the intermediate filament protein CP49 destabilises the lens fibre cell cytoskeleton and decreases lens optical quality, but does not induce cataract. Exp Eye Res 76(3):385–391

Sandilands A, Wang X, Hutcheson AM, James J, Prescott AR, Wegener A, Pekny M, Gong X, Quinlan RA (2004) Bfsp2 mutation found in mouse 129 strains causes the loss of CP49' and induces vimentin-dependent changes in the lens fibre cell cytoskeleton. Exp Eye Res 78(4):875–889. doi:10.1016/j.exer.2003.09.028

Sandilands A, Smith FJ, Lunny DP, Campbell LE, Davidson KM, MacCallum SF, Corden LD, Christie L, Fleming S, Lane EB, McLean WH (2013) Generation and characterisation of keratin 7 (K7) knockout mice. PLoS One 8(5):e64404. doi:10.1371/journal.pone.0064404

Santos M, Paramio JM, Bravo A, Ramirez A, Jorcano JL (2002) The expression of keratin k10 in the basal layer of the epidermis inhibits cell proliferation and prevents skin tumorigenesis. J Biol Chem 277(21):19122–19130. doi:10.1074/jbc.M201001200

Sasaki T, Gotow T, Shiozaki M, Sakaue F, Saito T, Julien JP, Uchiyama Y, Hisanaga S (2006) Aggregate formation and phosphorylation of neurofilament-L Pro22 Charcot-Marie-Tooth disease mutants. Hum Mol Genet 15(6):943–952. doi:10.1093/hmg/ddl011

Schwarz N, Windoffer R, Magin TM, Leube RE (2015) Dissection of keratin network formation, turnover and reorganization in living murine embryos. Sci Rep 5:9007. doi:10.1038/srep09007

Seltmann K, Fritsch AW, Kas JA, Magin TM (2013a) Keratins significantly contribute to cell stiffness and impact invasive behavior. Proc Natl Acad Sci U S A 110(46):18507–18512. doi:10.1073/pnas.1310493110

Seltmann K, Roth W, Kroger C, Loschke F, Lederer M, Huttelmaier S, Magin TM (2013b) Keratins mediate localization of hemidesmosomes and repress cell motility. J Invest Dermatol 133(1):181–190. doi:10.1038/jid.2012.256

Shah JV, Flanagan LA, Janmey PA, Leterrier JF (2000) Bidirectional translocation of neurofilaments along microtubules mediated in part by dynein/dynactin. Mol Biol Cell 11(10):3495–3508

Shah SB, Love JM, O'Neill A, Lovering RM, Bloch RJ (2012) Influences of desmin and keratin 19 on passive biomechanical properties of mouse skeletal muscle. J Biomed Biotechnol 2012:704061. doi:10.1155/2012/704061

Shea TB, Lee S (2011) Neurofilament phosphorylation regulates axonal transport by an indirect mechanism: a merging of opposing hypotheses. Cytoskelet (Hoboken) 68(11):589–595. doi:10.1002/cm.20535

Shrikant P, Benveniste EN (1996) The central nervous system as an immunocompetent organ: role of glial cells in antigen presentation. J Immunol 157(5):1819–1822

Sidhu GS, Chandra P, Cassai ND (2005) Merkel cells, normal and neoplastic: an update. Ultrastruct Pathol 29(3–4):287–294. doi:10.1080/01913120590951284

Sihag RK, Inagaki M, Yamaguchi T, Shea TB, Pant HC (2007) Role of phosphorylation on the structural dynamics and function of types III and IV intermediate filaments. Exp Cell Res 313(10):2098–2109. doi:10.1016/j.yexcr.2007.04.010

Simpson CL, Patel DM, Green KJ (2011) Deconstructing the skin: cytoarchitectural determinants of epidermal morphogenesis. Nat Rev Mol Cell Biol 12(9):565–580. doi:10.1038/nrm3175

Smith FJ, Porter RM, Corden LD, Lunny DP, Lane EB, McLean WH (2002) Cloning of human, murine, and marsupial keratin 7 and a survey of K7 expression in the mouse. Biochem Biophys Res Commun 297(4):818–827

Smith FJ, Irvine AD, Terron-Kwiatkowski A, Sandilands A, Campbell LE, Zhao Y, Liao H, Evans AT, Goudie DR, Lewis-Jones S, Arseculeratne G, Munro CS, Sergeant A, O'Regan G, Bale SJ, Compton JG, DiGiovanna JJ, Presland RB, Fleckman P, McLean WH (2006) Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. Nat Genet 38(3):337–342. doi:10.1038/ng1743

Snider NT, Omary MB (2014) Post-translational modifications of intermediate filament proteins: mechanisms and functions. Nat Rev Mol Cell Biol 15(3):163–177. doi:10.1038/nrm3753

Snider NT, Weerasinghe SV, Iniguez-Lluhi JA, Herrmann H, Omary MB (2011) Keratin hypersumoylation alters filament dynamics and is a marker for human liver disease and keratin mutation. J Biol Chem 286(3):2273–2284. doi:10.1074/jbc.M110.171314

Song S, Landsbury A, Dahm R, Liu Y, Zhang Q, Quinlan RA (2009) Functions of the intermediate filament cytoskeleton in the eye lens. J Clin Invest 119(7):1837–1848. doi:10.1172/JCI38277

Stone MR, O'Neill A, Lovering RM, Strong J, Resneck WG, Reed PW, Toivola DM, Ursitti JA, Omary MB, Bloch RJ (2007) Absence of keratin 19 in mice causes skeletal myopathy with mitochondrial and sarcolemmal reorganization. J Cell Sci 120(Pt 22):3999–4008. doi:10.1242/jcs.009241

Strnad P, Nuraldeen R, Guldiken N, Hartmann D, Mahajan V, Denk H, Haybaeck J (2013) Broad spectrum of hepatocyte inclusions in humans, animals, and experimental models. Compr Physiol 3(4):1393–1436. doi:10.1002/cphy.c120032

Szeverenyi I, Cassidy AJ, Chung CW, Lee BT, Common JE, Ogg SC, Chen H, Sim SY, Goh WL, Ng KW, Simpson JA, Chee LL, Eng GH, Li B, Lunny DP, Chuon D, Venkatesh A, Khoo KH, McLean WH, Lim YP, Lane EB (2008) The human intermediate filament database: comprehensive information on a gene family involved in many human diseases. Hum Mutat 29(3):351–360. doi:10.1002/humu.20652

Tamai Y, Ishikawa T, Bosl MR, Mori M, Nozaki M, Baribault H, Oshima RG, Taketo MM (2000) Cytokeratins 8 and 19 in the mouse placental development. J Cell Biol 151(3):563–572

Tanaka H, Goto H, Inoko A, Makihara H, Enomoto A, Horimoto K, Matsuyama M, Kurita K, Izawa I, Inagaki M (2015) Cytokinetic failure-induced tetraploidy develops into aneuploidy, triggering skin aging in phosphovimentin-deficient mice. J Biol Chem 290(21):12984–12998. doi:10.1074/jbc.M114.633891

Tao GZ, Looi KS, Toivola DM, Strnad P, Zhou Q, Liao J, Wei Y, Habtezion A, Omary MB (2009) Keratins modulate the shape and function of hepatocyte mitochondria: a mechanism for protection from apoptosis. J Cell Sci 122(Pt 21):3851–3855. doi:10.1242/jcs.051862

Toivola DM, Omary MB, Ku NO, Peltola O, Baribault H, Eriksson JE (1998) Protein phosphatase inhibition in normal and keratin 8/18 assembly-incompetent mouse strains supports a functional role of keratin intermediate filaments in preserving hepatocyte integrity. Hepatology 28(1):116–128. doi:10.1002/hep.510280117

Toivola DM, Nieminen MI, Hesse M, He T, Baribault H, Magin TM, Omary MB, Eriksson JE (2001) Disturbances in hepatic cell-cycle regulation in mice with assembly-deficient keratins 8/18. Hepatology 34(6):1174–1183. doi:10.1053/jhep.2001.29374

Toivola DM, Krishnan S, Binder HJ, Singh SK, Omary MB (2004) Keratins modulate colonocyte electrolyte transport via protein mistargeting. J Cell Biol 164(6):911–921. doi:10.1083/jcb.200308103

Toivola DM, Nakamichi I, Strnad P, Michie SA, Ghori N, Harada M, Zeh K, Oshima RG, Baribault H, Omary MB (2008) Keratin overexpression levels correlate with the extent of spontaneous pancreatic injury. Am J Pathol 172(4):882–892. doi:10.2353/ajpath.2008.070830

Toivola DM, Habtezion A, Misiorek JO, Zhang L, Nystrom JH, Sharpe O, Robinson WH, Kwan R, Omary MB (2015) Absence of keratin 8 or 18 promotes antimitochondrial autoantibody formation in aging male mice. FASEB J 29(12):5081–5089. doi:10.1096/fj.14-269795

Tong X, Coulombe PA (2006) Keratin 17 modulates hair follicle cycling in a TNFalpha-dependent fashion. Genes Dev 20(10):1353–1364. doi:10.1101/gad.1387406

Tsuruta D, Hashimoto T, Hamill KJ, Jones JC (2011) Hemidesmosomes and focal contact proteins: functions and cross-talk in keratinocytes, bullous diseases and wound healing. J Dermatol Sci 62(1):1–7. doi:10.1016/j.jdermsci.2011.01.005

Tumbar T, Guasch G, Greco V, Blanpain C, Lowry WE, Rendl M, Fuchs E (2004) Defining the epithelial stem cell niche in skin. Science 303(5656):359–363. doi:10.1126/science.1092436

Vardjan N, Gabrijel M, Potokar M, Svajger U, Kreft M, Jeras M, de Pablo Y, Faiz M, Pekny M, Zorec R (2012) IFN-gamma-induced increase in the mobility of MHC class II compartments in astrocytes depends on intermediate filaments. J Neuroinflammation 9:144. doi:10.1186/1742-2094-9-144

Vijayaraj P, Kroger C, Reuter U, Windoffer R, Leube RE, Magin TM (2009) Keratins regulate protein biosynthesis through localization of GLUT1 and -3 upstream of AMP kinase and Raptor. J Cell Biol 187(2):175–184. doi:10.1083/jcb.200906094

Vijayaraj P, Kroeger C, Reuter U, Hartmann D, Magin TM (2010) Keratins regulate yolk sac hematopoiesis and vasculogenesis through reduced BMP-4 signaling. Eur J Cell Biol 89(4):299–306. doi:10.1016/j.ejcb.2009.10.016

Wallace L, Roberts-Thompson L, Reichelt J (2012) Deletion of K1/K10 does not impair epidermal stratification but affects desmosomal structure and nuclear integrity. J Cell Sci 125(Pt 7):1750–1758. doi:10.1242/jcs.097139

Wang L, Ho CL, Sun D, Liem RK, Brown A (2000) Rapid movement of axonal neurofilaments interrupted by prolonged pauses. Nat Cell Biol 2(3):137–141. doi:10.1038/35004008

Wang YF, Lang HY, Yuan J, Wang J, Wang R, Zhang XH, Zhang J, Zhao T, Li YR, Liu JY, Zeng LH, Guo GZ (2013) Overexpression of keratin 17 is associated with poor prognosis in epithelial ovarian cancer. Tumour Biol 34(3):1685–1689. doi:10.1007/s13277-013-0703-5

Watt FM (1998) Epidermal stem cells: markers, patterning and the control of stem cell fate. Philos Trans R Soc Lond B Biol Sci 353(1370):831–837. doi:10.1098/rstb.1998.0247

Weerasinghe SV, Ku NO, Altshuler PJ, Kwan R, Omary MB (2014) Mutation of caspase-digestion sites in keratin 18 interferes with filament reorganization, and predisposes to hepatocyte necrosis and loss of membrane integrity. J Cell Sci 127(Pt 7):1464–1475. doi:10.1242/jcs.138479

Weisleder N, Soumaka E, Abbasi S, Taegtmeyer H, Capetanaki Y (2004a) Cardiomyocyte-specific desmin rescue of desmin null cardiomyopathy excludes vascular involvement. J Mol Cell Cardiol 36(1):121–128

Weisleder N, Taffet GE, Capetanaki Y (2004b) Bcl-2 overexpression corrects mitochondrial defects and ameliorates inherited desmin null cardiomyopathy. Proc Natl Acad Sci U S A 101(3):769–774. doi:10.1073/pnas.0303202101

Widestrand A, Faijerson J, Wilhelmsson U, Smith PL, Li L, Sihlbom C, Eriksson PS, Pekny M (2007) Increased neurogenesis and astrogenesis from neural progenitor cells grafted in the hippocampus of GFAP−/− Vim−/− mice. Stem Cells 25(10):2619–2627. doi:10.1634/stemcells.2007-0122

Wilhelmsson U, Faiz M, de Pablo Y, Sjoqvist M, Andersson D, Widestrand A, Potokar M, Stenovec M, Smith PL, Shinjyo N, Pekny T, Zorec R, Stahlberg A, Pekna M, Sahlgren C, Pekny M (2012) Astrocytes negatively regulate neurogenesis through the Jagged1-mediated Notch pathway. Stem Cells 30(10):2320–2329. doi:10.1002/stem.1196

Windoffer R, Beil M, Magin TM, Leube RE (2011) Cytoskeleton in motion: the dynamics of keratin intermediate filaments in epithelia. J Cell Biol 194(5):669–678. doi:10.1083/jcb.201008095

Winter L, Wiche G (2013) The many faces of plectin and plectinopathies: pathology and mechanisms. Acta Neuropathol 125(1):77–93. doi:10.1007/s00401-012-1026-0

Winter L, Abrahamsberg C, Wiche G (2008) Plectin isoform 1b mediates mitochondrion-intermediate filament network linkage and controls organelle shape. J Cell Biol 181(6):903–911. doi:10.1083/jcb.200710151

Winter L, Kuznetsov AV, Grimm M, Zeold A, Fischer I, Wiche G (2015) Plectin isoform P1b and P1d deficiencies differentially affect mitochondrial morphology and function in skeletal muscle. Hum Mol Genet 24(16):4530–4544. doi:10.1093/hmg/ddv184

Wojcik SM, Bundman DS, Roop DR (2000) Delayed wound healing in keratin 6a knockout mice. Mol Cell Biol 20(14):5248–5255

Wong P, Coulombe PA (2003) Loss of keratin 6 (K6) proteins reveals a function for intermediate filaments during wound repair. J Cell Biol 163(2):327–337. doi:10.1083/jcb.200305032

Woo WM, Goncharov A, Jin Y, Chisholm AD (2004) Intermediate filaments are required for *C. elegans* epidermal elongation. Dev Biol 267(1):216–229. doi:10.1016/j.ydbio.2003.11.007

Xu Z, Marszalek JR, Lee MK, Wong PC, Folmer J, TO C, Hsieh ST, Griffin JW, Cleveland DW (1996) Subunit composition of neurofilaments specifies axonal diameter. J Cell Biol 133(5):1061–1069

Yamaguchi T, Goto H, Yokoyama T, Sillje H, Hanisch A, Uldschmid A, Takai Y, Oguri T, Nigg EA, Inagaki M (2005) Phosphorylation by Cdk1 induces Plk1-mediated vimentin phosphorylation during mitosis. J Cell Biol 171(3):431–436. doi:10.1083/jcb.200504091

Yang Y, Dowling J, Yu QC, Kouklis P, Cleveland DW, Fuchs E (1996) An essential cytoskeletal linker protein connecting actin microfilaments to intermediate filaments. Cell 86(4):655–665

Yasui Y, Amano M, Nagata K, Inagaki N, Nakamura H, Saya H, Kaibuchi K, Inagaki M (1998) Roles of Rho-associated kinase in cytokinesis; mutations in Rho-associated kinase phosphorylation sites impair cytokinetic segregation of glial filaments. J Cell Biol 143(5):1249–1258

Yasui Y, Goto H, Matsui S, Manser E, Lim L, Nagata K, Inagaki M (2001) Protein kinases required for segregation of vimentin filaments in mitotic process. Oncogene 20(23):2868–2876. doi:10.1038/sj.onc.1204407

Yatsunami J, Fujiki H, Suganuma M, Yoshizawa S, Eriksson JE, Olson MO, Goldman RD (1991) Vimentin is hyperphosphorylated in primary human fibroblasts treated with okadaic acid. Biochem Biophys Res Commun 177(3):1165–1170

Yin X, TO C, Griffin JW, Tu P, Lee VM, Li C, Roder J, Trapp BD (1998) Myelin-associated glycoprotein is a myelin signal that modulates the caliber of myelinated axons. J Neurosci 18(6):1953–1962

Yuan A, Nixon RA, Rao MV (2006) Deleting the phosphorylated tail domain of the neurofilament heavy subunit does not alter neurofilament transport rate in vivo. Neurosci Lett 393(2–3):264–268. doi:10.1016/j.neulet.2005.10.029

Yuan A, Sasaki T, Rao MV, Kumar A, Kanumuri V, Dunlop DS, Liem RK, Nixon RA (2009) Neurofilaments form a highly stable stationary cytoskeleton after reaching a critical level in axons. J Neurosci 29(36):11316–11329. doi:10.1523/jneurosci.1942-09.2009

Yuan A, Sershen H, Veeranna BBS, Kumar A, Hashim A, Berg M, Lee JH, Sato Y, Rao MV, Mohan PS, Dyakin V, Julien JP, Lee VM, Nixon RA (2015) Neurofilament subunits are integral components of synapses and modulate neurotransmission and behavior in vivo. Mol Psychiatry 20(8):986–994. doi:10.1038/mp.2015.45

Zatloukal K, Stumptner C, Lehner M, Denk H, Baribault H, Eshkind LG, Franke WW (2000) Cytokeratin 8 protects from hepatotoxicity, and its ratio to cytokeratin 18 determines the ability of hepatocytes to form Mallory bodies. Am J Pathol 156(4):1263–1274. doi:10.1016/S0002-9440(10)64997-8

Zatloukal K, French SW, Stumptner C, Strnad P, Harada M, Toivola DM, Cadrin M, Omary MB (2007) From Mallory to Mallory-Denk bodies: what, how and why? Exp Cell Res 313(10):2033–2049. doi:10.1016/j.yexcr.2007.04.024

Zhang J, Bang ML, Gokhin DS, Lu Y, Cui L, Li X, Gu Y, Dalton ND, Scimia MC, Peterson KL, Lieber RL, Chen J (2008) Syncoilin is required for generating maximum isometric stress in skeletal muscle but dispensable for muscle cytoarchitecture. Am J Physiol Cell Physiol 294(5):C1175–C1182. doi:10.1152/ajpcell.00049.2008

Zhang H, Landmann F, Zahreddine H, Rodriguez D, Koch M, Labouesse M (2011) A tension-induced mechanotransduction pathway promotes epithelial morphogenesis. Nature 471(7336):99–103. doi:10.1038/nature09765

Zhou Q, Toivola DM, Feng N, Greenberg HB, Franke WW, Omary MB (2003) Keratin 20 helps maintain intermediate filament organization in intestinal epithelia. Mol Biol Cell 14(7):2959–2971. doi:10.1091/mbc.E03-02-0059

Zhu Q, Couillard-Despres S, Julien JP (1997) Delayed maturation of regenerating myelinated axons in mice lacking neurofilaments. Exp Neurol 148(1):299–316. doi:10.1006/exnr.1997.6654

Zhu Q, Lindenbaum M, Levavasseur F, Jacomy H, Julien JP (1998) Disruption of the NF-H gene increases axonal microtubule content and velocity of neurofilament transport: relief of axonopathy resulting from the toxin beta,beta'-iminodipropionitrile. J Cell Biol 143(1):183–193

Chapter 8
Filamentous Structure of Hard β-Keratins in the Epidermal Appendages of Birds and Reptiles

R.D. Bruce Fraser and David A.D. Parry

Contents

Abstract The structures of avian and reptilian epidermal appendages, such as feathers, claws and scales, have been modelled using X-ray diffraction and electron microscopy data, combined with sequence analyses. In most cases, a family of closely related molecules makes up the bulk of the appendage, and each of these molecules contains a central β-rich 34-residue segment, which has been identified

R.D.B. Fraser
Institute of Fundamental Sciences, Massey University,
Private Bag 11-222, Palmerston North 4442, New Zealand

28 Satinay Drive, Noosa Parklands, Tewantin, Qld 4565, Australia

D.A.D. Parry (✉)
Institute of Fundamental Sciences and Riddet Institute, Massey University,
Private Bag 11-222, Palmerston North 4442, New Zealand
e-mail: d.parry@massey.ac.nz

as the principal component of the framework of the 3.4 nm diameter filaments. The N- and C-terminal segments form the matrix component of the filament/matrix complex. The 34-residue β-rich central domains occur in pairs, related by either a parallel dyad or a perpendicular dyad axis, and form a β-sandwich stabilized by apolar interactions. They are also twisted in a right-handed manner. In feather, the filaments are packed into small sheets and it is possible to determine their likely orientation within the sheets from the low-angle X-ray diffraction data. The physical properties of the various epidermal appendages can be related to the amino acid sequence and composition of defined molecular segments characteristic of the chains concerned.

Keywords Squamates • Archosaurs • Feather keratin X-ray pattern • Filament packing in feather • Mechanical properties of β-keratins • High-glycine-tyrosine proteins

8.1   Introduction

There is strong evidence, from electron microscopy studies (Fig. 8.1), that avian and reptilian hard keratins, such as feathers and claws, have a composite structure consisting of filaments embedded in a matrix (Filshie and Rogers 1962; Alexander 1970; Maderson et al. 1972; Landmann 1979; Alibardi et al. 2006, 2009). This type of composite gives a high tensile strength and a degree of flexibility that depends on the composition of the matrix.

A common feature of the hard β-keratin proteins isolated from feathers, claws, scales and other avian and reptilian hard keratins is that they contain segments associated with both the filaments and the matrix in a single chain. This is in sharp contrast to the mammalian keratins where the filaments and the matrix are formed from different types of protein, with in excess of 100 unique proteins in any particular appendage. It is important to note that, in addition to hard β-keratin, many avian and reptilian appendages contain intermediate filaments (IF), usually in discrete layers, that are similar to those found in mammalian keratins (Rudall 1947; Dalla Valle et al. 2005, 2007a, b, 2009a, b, 2010; Hallahan et al. 2009).

Over the past 20 years there has been a rapid increase in the number of amino acid sequences of β-keratins deposited in the protein databases (see, for example, those from emu (O'Donnell 1973), silver gull (O'Donnell and Inglis 1974), chicken scale and feather (Gregg et al. 1984), lizard claw (Inglis et al. 1987), snake (Dalla Valle et al. 2007a), gecko (Dalla Valle et al. 2007b), turtle (Dalla Valle et al. 2009a), crocodile (Dalla Valle et al. 2009b), green anole lizard (Dalla Valle et al. 2010), and chicken and zebra finch (Greenwold and Sawyer 2010). These data and those from other species not listed above now number several hundred, and their study has provided new insights into the details of the structure of the filament-matrix composite. To put the sequence data into context, phylogenetic analysis has shown that

Fig. 8.1 Cross-section of feather keratin showing filaments ca. 3 nm in diameter (mf) embedded in a matrix (Filshie and Rogers 1962). The filaments tend to aggregate in sheets (l), which leads to the unusual X-ray diffraction pattern reported earlier by Bear and Rugo (1951) (Reprinted from Fraser and Parry 2011a with permission from Elsevier)
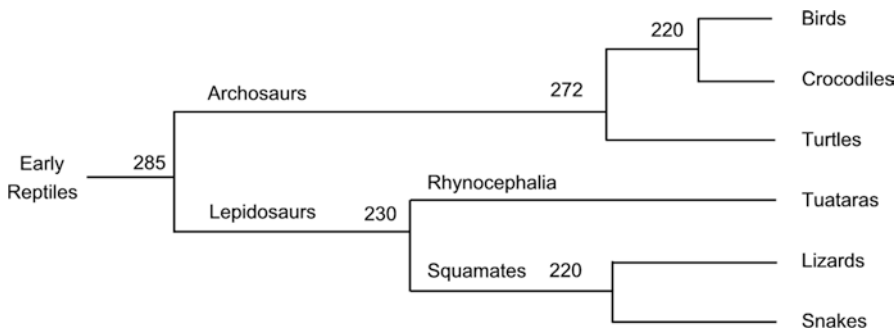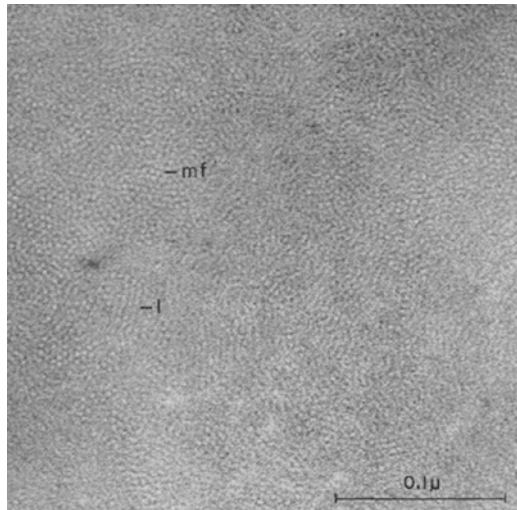
Fig. 8.2 A phylogenetic classification of birds and reptiles with estimates of the branching times measured in millions of years. The figures are based on studies of mitochondrial DNA and are taken from the work of Rest et al. (2003) (Reprinted from Fraser and Parry 2014 with permission from Elsevier)

the birds and reptiles (the Sauropsids) can be subdivided into Archosaurs (birds, crocodiles and turtles) and Lepidosaurs. The latter can, in turn, be subdivided into the Squamates (snakes and lizards) and the Rhynocephalia (tuataras) (Fig. 8.2). Similarities and differences between the sequences of the Archosaurs and the Squamates have proved to be particularly informative, and these are discussed in Sect. 8.2.

X-ray diffraction patterns obtained from well-oriented specimens of hard β-keratins, such as feather rachis, have revealed a highly ordered structure, presumably originating primarily from the filamentous portion of the structure. Since the pattern has similarities to that derived from stretched α-keratin, it is believed that the framework of the filaments in the avian and reptilian keratins is based on the β-sheet conformation rather than the α-helical conformation (Astbury and Marwick 1932).

Subsequent X-ray studies (Schorr and Krimm 1961; Fraser and MacRae 1959, 1963) further indicated that the filaments consist of helical arrays of small β-crystallites. When the first amino acid sequence of a β-keratin was completed (emu feather), two additional and confirmatory elements of the structure became apparent. Firstly, an enzymatic fragment of this sequence was shown by infrared spectroscopy to be rich in β-structure (Suzuki 1973). Secondly, in a detailed study of the sequence, Fraser and MacRae (1976) investigated the distribution of β-favouring residues and found that they were concentrated in a central domain. In addition, using Fourier transform techniques, it was found that there was a periodicity of eight-residues in the β-propensity within this domain that was π out of phase with the turn-propensity, suggesting that the chain looped back and forth. The predicted length of the loops was 2.4 nm and this precisely matched the spacing of the dominant meridional reflection. Strong support for looped chains was provided by studies of the infrared spectrum (Fraser and Suzuki 1965), which revealed that the β-sheets were primarily in the antiparallel rather than the parallel conformation. When a sequence for a reptilian appendage (goanna claw) became available (Inglis et al. 1987) it was found that a similar central domain was present (Fraser and Parry 1996). Since then it has been shown that this β-containing central domain is conserved across all species and appendages (Fraser and Parry 2008, 2011a, 2014). Details of the conformation of the polypeptide chain in the β-crystallites, derived from the X-ray fibre diffraction data are described in Sect 8.3.

Since the structure of the filament is not determinable at atomic resolution from the available X-ray data, it is necessary to use computer modelling to derive insights into aspects of the three-dimensional conformation of the filament. Using this approach, it can be shown that the β-crystallites, through a highly specific lateral and axial polymerization process, form the framework of the filaments. Furthermore, the particular amino acids involved in both the assembly and stabilization of the component β-sheets in the dimer and in the axial aggregation of those dimers can be recognized, and this is described in Sect. 8.4.

In the case of feather keratin there is evidence in both the low-angle X-ray diffraction patterns and in the electron micrographs of a two-dimensional latticework of filaments. These data are indicative of small sheets containing approximately 4–10 filaments, each of which is organized identically with the same azimuth within the sheet. The detailed interpretation of these data is discussed in Sect. 8.5.

The matrix of the β-keratins is composed of the bulk of the N- and C-terminal domains of the particular protein chain involved. These sequences have a wide variety of lengths and chemical characteristics, and include various sequence repeats. The terminal domains display both conserved and variable features (Fraser and Parry 2011a, 2014) and, in many cases, these can be associated with particular mechanical and physical properties (Fraser and Parry 2014). In particular, there are regions that are glycine-tyrosine-rich (HGT), charged- and cysteine-rich, and also glycine-rich. In Sect. 8.6 the likely physical and mechanical attributes of such regions are discussed.

8.2   Sequence Similarities and Differences

Because the lengths of the sequences vary considerably between species and between epidermal appendages, it is useful for the sequence analysis work to define a reference point in the sequence that is common to all chains, i.e. a residue that is highly conserved and which thereby allows ready comparisons of position between chains to be made. A PGP motif is the most highly conserved sequence in all β-keratin sequences and for this reason the glycine G was chosen as the origin, i.e. $n_G = 0$ (Fraser and Parry 2014). This nomenclature will be used in the remainder of this chapter.

By comparing representative groups of all the available β-keratin sequences (Figs. 8.3, 8.4 and 8.5) a number of general features have become evident (Fraser and Parry 2014). Firstly, there is a central segment 34-residues in length that is highly homologous across all of the chains ($n_G = -21$ to 12, see Sect. 8.2.1). Secondly, all of the avian feather sequences, excluding emu feather, contain an extended homologous region ($n_G = -35$ to 20) that includes the 34-residue segment. The two one-residue insertions in emu feather ($n_G = -21$ and 13), which represents the only ratite in the sequence databases, occur at the ends of the 34-residue segment. Thirdly, the sequences of the Archosaurs as a group and of the Squamates as a group each display many similar features in their N- and C-terminal domains as discussed in Sects. 8.2.2 and 8.2.3.

8.2.1   Filament-Forming Central Domain

Sawyer et al. (2000) noted the very high homology that existed between family members of the avian β-keratins. They also reported a 20-residue segment, situated in the central domains of avian sequences, that was closely homologous to that found in alligator claw ($n_G = -15$ to 4). Alibardi and Toni (2007) showed that this same segment was also present in the Archosaurs and the Squamates, and Dalla Valle et al. (2010) proposed that it had a role in filament assembly. This 20-residue segment forms part of the looped central domain identified by Fraser and MacRae (1976), and later refined to the present central domain. A detailed analysis of the infrared data on seagull feather rachis was consistent with an antiparallel β-conformation and a length of around 28 residues. Allowing for two residues in each of the hairpin bends this compares favourably with the value of 26 residues in the detailed model for the molecule discussed later in this Section.

Fraser and Parry (2011a) designed a number of structural probes in an attempt to assess more precisely the β-containing region in a wide selection of chains across both species and appendages. These probes were based on:

(a) The β- and turn- propensities to assess those portions of the sequences with the greatest propensity for these elements of secondary structure
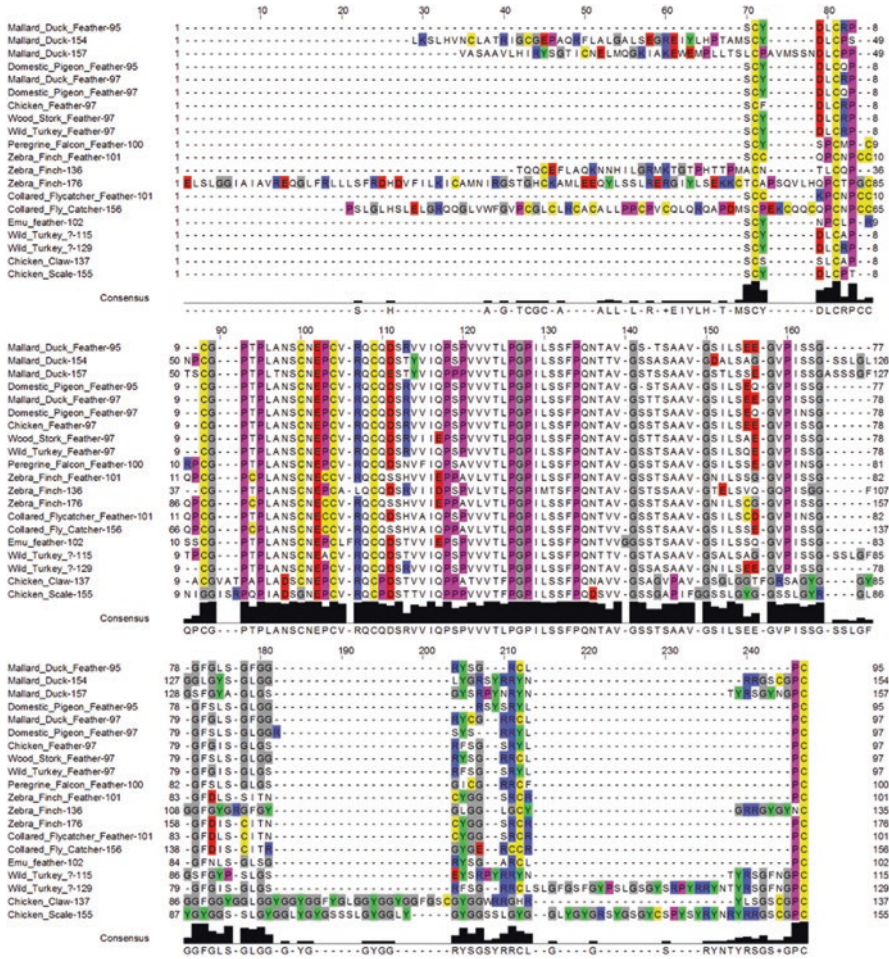
Fig. 8.3 A comparison of the sequences of a selection of the avian Archosaurs. The feather keratins are closely homologous over their entire lengths. However, claw and scale keratins contain glycine-rich insertions in their C-terminal domains. In some sequences too there are extensions at the extreme N-terminal end of the sequences that are rich in charged and cysteine residues. The colour scheme used is as follows: cysteine (*yellow*), proline (*magenta*), glycine (*grey*), tyrosine (*green*), acidic residues (*red*) and basic residues (*blue*) (Reprinted from Fraser and Parry 2014 with permission from Elsevier)

(b) Hydropathy indices to recognize regions with a two-residue period expected of a β-sheet with one side predominantly apolar

(c) Homology comparisons to ascertain those segments most highly conserved across all sequences

(d) Fourier transform analyses to define regions of regularity in the conformational parameters.
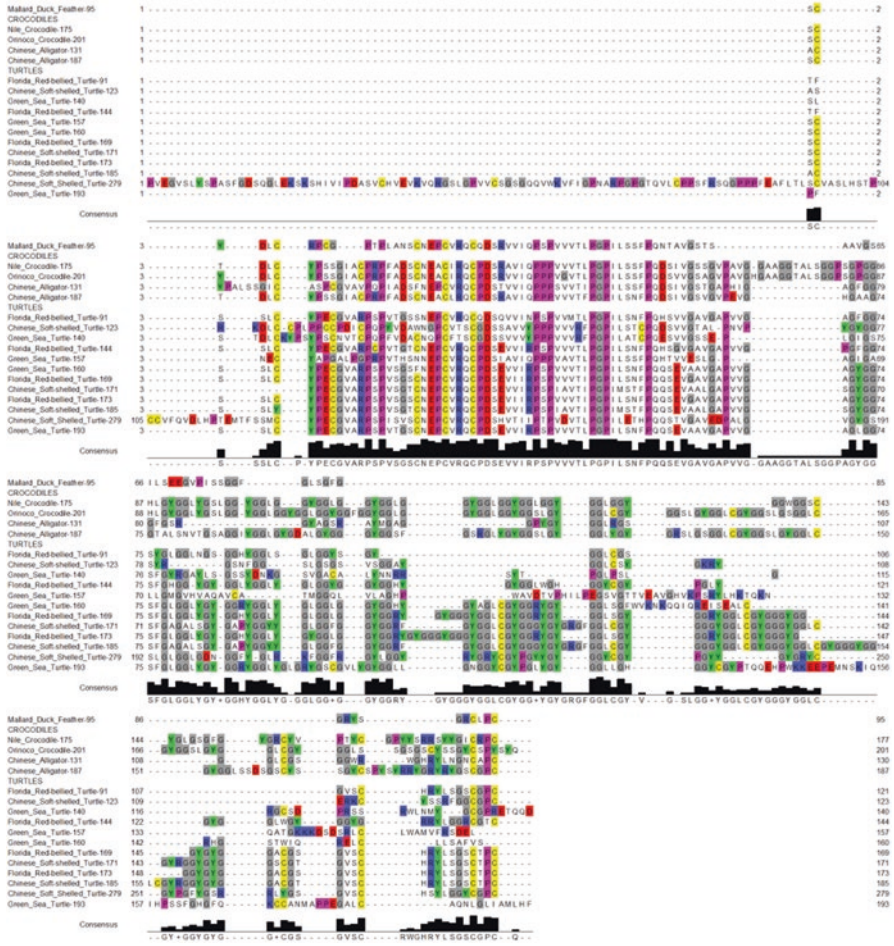
Fig. 8.4 A comparison of the sequences of a selection of the reptilian Archosaurs. When compared to feather each of the sequences contains large insertions of segments rich in glycine and tyrosine residues (HGT) in their C-terminal domains. There is also an example of a large extension at the N-terminal end of the sequence of Chinese soft-shelled turtle that is rich in charged and cysteine residues. The colour scheme used is the same as that defined in Fig. 8.3 (Reprinted from Fraser and Parry 2014 with permission from Elsevier)

A consistent answer emerged in which a homologous 34-residue β-favouring sequence was recognized in all sequences ($n_G = -21$ to 12). Furthermore, it was shown that one face of the putative antiparallel β-sheet thus formed comprised predominantly apolar residues. In these studies it was also noted that there could be small departures from the idealized sheet structure shown in Fig. 6 involving both the number of residues in each strand and the number of residues involved in the turns. Independent modelling of this region by Calvaresi et al. (2016), using an *ab initio* protein folding and structure prediction algorithm, confirmed the presence of
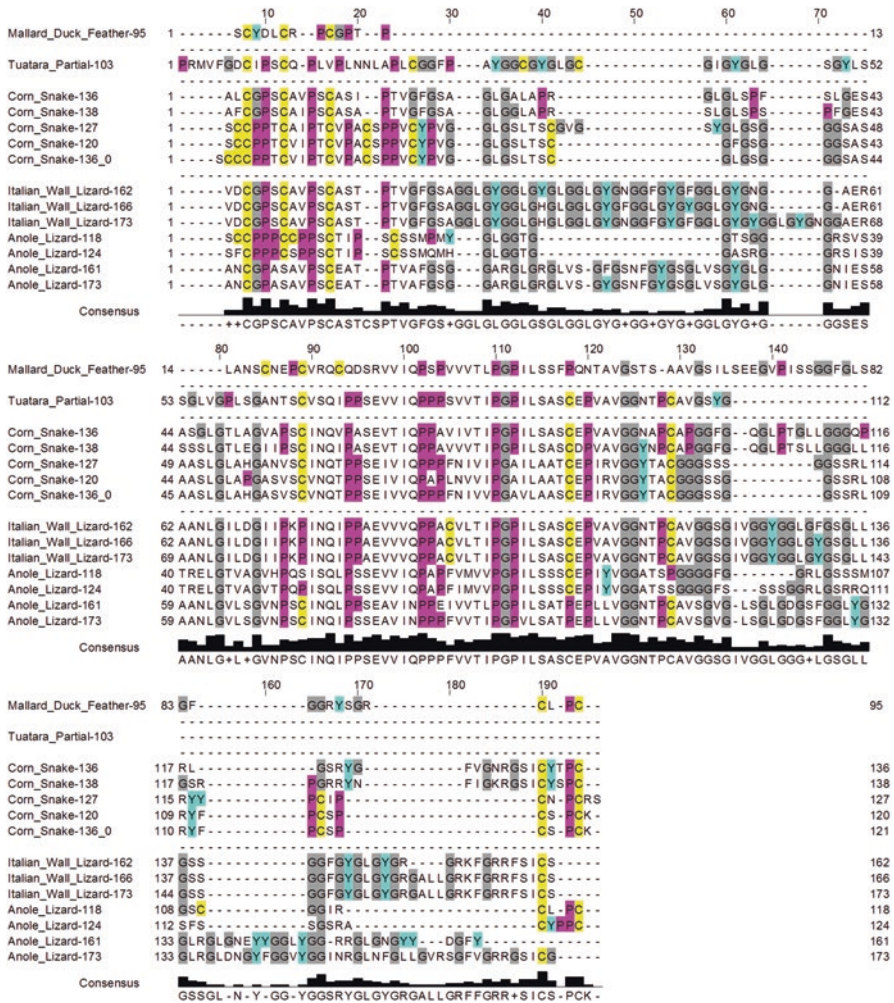
Fig. 8.5 An alignment of a selection of amino acid sequences from the Squamates. Compared with feather all the sequences have insertions of HGT segments in both the N-terminal and the C-terminal domains. A partial sequence from a tuatara (Rhynocephalia) is also included. The colour scheme used is the same as that defined in Fig. 8.3 (Reprinted from Fraser and Parry 2014 with permission from Elsevier)

a 34-residue β-rich central domain composed of antiparallel strands separated by tight turns (Chou 2000). In addition, molecular mechanics/Poisson Boltzmann surface area analysis confirmed the amphipathic nature of one face of the β-sheet.

On the basis of their study of avian and reptilian sequences it was proposed (Fraser and Parry 2014) that the β-crystallite contained five strands (three complete central strands and two outer partial ones) connected by four bends that consist of either two or four residues depending on the type of bend (Fig. 8.6). It was concluded
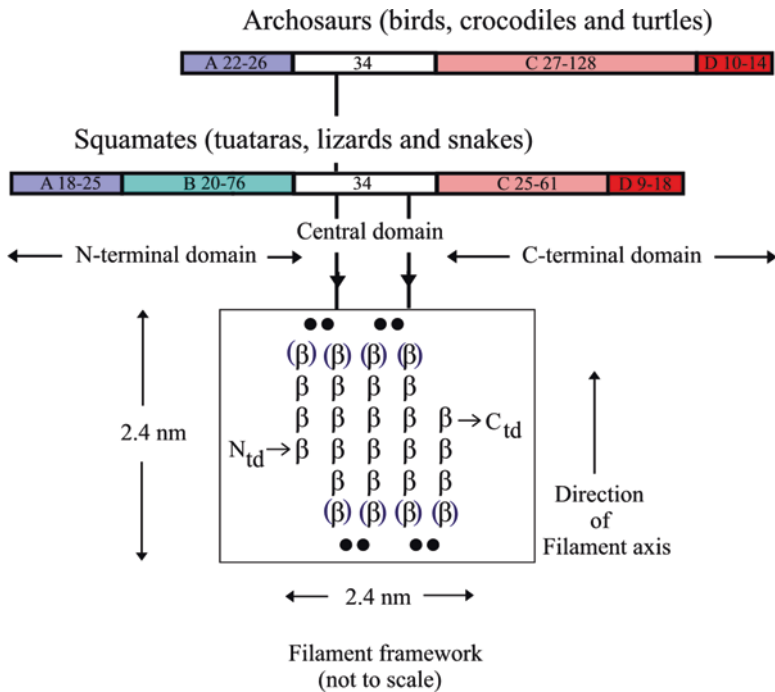
Fig. 8.6 A summary of the main findings from sequence studies. The Archosaurs and Squamates have a common central domain consisting of 34-residues, which forms an antiparallel β-sheet with tight turns. The exact lengths of the strands and locations and precise nature of the turns are unknown, but are probably less regular than depicted here (Fraser and Parry 2011a)

(Fraser and Parry 2011a, 2014) that the central domain present in all of the Archosaurs and Squamates must be associated with the framework of the filament. Apart from short regions, 7 to 11 residues in length, immediately N- and C-terminal to the 34-residue segment ($n_G = -28$ to -22 and $n_G = 13$ to 23), that have been identified as likely to also play some role in filament assembly (Fraser and Parry 2014), the N- and C-terminal domains form the bulk of the matrix.

8.2.2 N-Terminal Domain

The N-terminal domain is that part of the chain lying N-terminal to $n_G = -21$ i.e. from the N-terminus through to and including $n_G = -22$. From a study of a wide selection of the available β-keratin sequences from the Archosaurs and the Squamates it is apparent that significant differences exist between them (Fig. 8.6). Almost all of the Archosaurian N-terminal sequences are short and of similar length (22–26 residues) and, furthermore, are homologous to one another (Fraser and Parry 2011a). This is indicative of a common structure and function. The N-terminal

domains of the Squamate sequences, in contrast to those of the Archosaurs, are much longer and more variable in length (range 7–94 residues), but they can be divided into two sub-domains (A and B) on the basis of amino acid composition. Subdomain A is poor in glycine and aromatic residues but relatively rich in cysteine, proline and serine residues. In contrast, subdomain B is particularly rich in glycine and tyrosine residues and is sometimes referred to as a HGT segment (high glycine-tyrosine) by analogy with those proteins found in the matrix of the hard α-keratins. Subdomain A is most commonly about 18 residues in length though some variations up to a value of 25 residues have been observed. These sequences also display some homology to one another, but not to those seen in the N-terminal domain of the Archosaurs (Fraser and Parry 2011a). Most of the subdomain B sequences consist of repeats, ranging from two to seven residues in length, largely based on glycine, serine, alanine, leucine and aromatic residues. In addition to the features listed above that are specific to the Archosaurs or the Squamates there are a few higher molecular weight avian and reptilian keratins which have sequences that extend beyond the "normal" SCY N-terminus, and these extensions are rich in charged, cysteine and glycine residues.

8.2.3   C-Terminal Domain

The C-terminal domain is represented by all of the chain lying C-terminal to $n_G = 12$ (i.e. from and including $n_G = 13$ through to the C-terminus). The lengths of these domains vary considerably within the Archosaurs (41–142 residues) and within the Squamates (20–70 residues). It is evident, nonetheless, that all the Archosaur and Squamate sequences (other than those from feather) can be divided into two subdomains (C and D) on the basis of amino acid composition (Fraser and Parry 2011a). Subdomain C (27–128 and 25–61 residues long in the Archosaurs and Squamates respectively) is rich in glycine, serine and tyrosine residues (HGT segments), and often contains sequence repeats. These repeats can be as short as two- or three-residues, but may be as long as 13 residues, as in chicken scale. The repeats differ, however, from those seen in subdomain B in the N-terminal domain. Subdomain D in the Archosaurs (10–14 residues long) and the Squamates (9–18 residues long) is characterized by elevated levels of cysteine and often arginine residues (Fraser and Parry 2011a). In feather keratin, however, the C-terminal domain contains no cysteine apart from the terminal residue, and it follows that this region, in contrast to the N-terminal domain, can play only a very minor role in the formation of inter- or intra-filament disulphide bonds.

Two other observations on the C-terminal domain have been made. First, the Archosaurian chains, as well as many of those in the Squamates, are terminated by a PC motif. Its role, however, is currently unknown. Second, the feather sequences and those from Nile and Orinoco crocodiles contain small glycine-rich segments ($n_G = 32$–45 and $n_G = 23$ to 36 respectively).
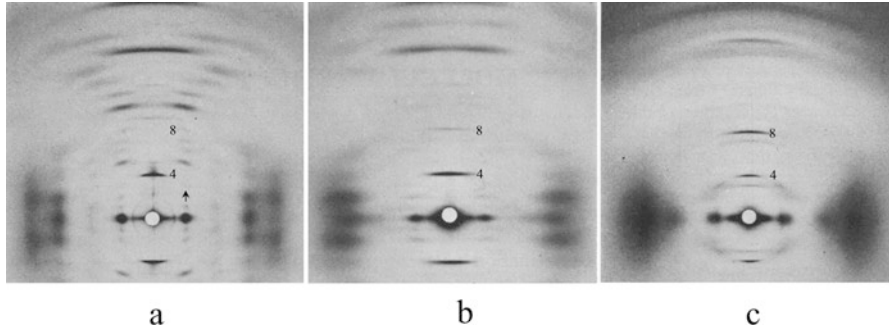
Fig. 8.7. X-ray diffraction patterns obtained from (**a**) seagull feather rachis (*Larus novae-hollandiae*), (**b**) the same specimen after pressing in steam, (**c**) goanna claw *Varanus varius*) (Reprinted from Fraser and Parry 2011b with permission from Elsevier)

8.3   Information on the Molecular Structure Derived from X-Ray Diffraction Studies

The X-ray fibre diffraction pattern of seagull feather rachis, shown in Fig. 8.7a, is a mixture of diffraction from the filaments plus ordered matrix, and an interference pattern associated with the packing of the filaments into sheets. A technique was developed for isolating the contribution of the filamentous component (Fraser and MacRae (1963) and the result is shown in Fig. 8.7b.

The diffraction pattern in Fig. 8.7b is typical of a basic helix (Klug et al. 1958) with a unit height of 2.4 nm and a pitch of |9.6| nm (i.e. four units per turn), and Fraser and MacRae (1963) suggested that the filaments were helical arrays of small β-crystallite units. The discrete row line reflections in Fig. 8.7a stem from the fact that in feather rachis the filaments are packed into sheets with a lateral spacing of 3.4 nm (Bear and Rugo 1951). Of all the epidermal appendages studied thus far it is only feather rachis that shows well-developed row lines. The differences between the intensities of the meridional reflections with indices 4 and 8 in Fig. 8.7a, b suggests that portions of the matrix are ordered when the filaments are packed into sheets but that this order is destroyed during the pressing in steam.

Minor differences in the pitch (*P*) of the filament helix are found in different species of avian and reptilian keratins: chicken scale (*P* = −9.2 nm, Stewart 1977), snake scale (*P* = −9.28 nm; Rudall 1947) and lizard claw (*P* = −9.85 nm; Rudall 1947). The quality of the diffraction patterns, however, are markedly inferior to those obtained from feather. Although small but significant differences in the value of *P* are evident there is every indication that the overall fourfold screw symmetry of the filaments is conserved. The variation in helix parameters undoubtedly relate to the differences in sequences and lengths of the chains in the various structures. Considerable differences in the distribution of intensities are also observed across

species as, for example, between seagull feather in Fig. 8.7a and goanna claw in Fig. 8.7c. These too are likely to be a consequence of the differences in amino acid composition in the N- and C- terminal domains.

The diffraction pattern in Fig. 8.7b was investigated in detail by Fraser et al. (1971) with the object of producing a detailed model for the framework of the filament. Earlier studies of the infrared spectrum of seagull feather (Fraser and Suzuki 1965) had established that the β-sheets had an antiparallel chain conformation, and the initial model consisted of a helical array of pairs of such sheets. It soon became clear that the sheets were twisted and the final model is shown in Fig. 8.8. Subsequent studies suggested that the sheet consisted of three complete central strands and two partial outer strands linked together by hairpin bends (Fraser and Parry 2008, 2014). This model embraces three possible conformations, depending on the interrelationship between the two halves of the β-sandwich. If they are non-identical (Model 1) the line group (Klug et al. 1958) is **s** with a unit height h = 2.4 nm, a unit twist $|t|$ = 90° and a pitch $|P|$ = 9.6 nm. In more likely event that the conformations are identical and the two halves of the sandwich are related by a dyad axis, the line group depends on the orientation of the dyad. If it is parallel to the fibre axis the structure (Model 2) will belong to line group **sr** (*N=2*) and comprise a pair of similarly-oriented basic helices with unit heights h = 2.4 nm, unit twists $|t|$ = 45° and pitches $|P|$ = 19.2 nm. The parallel dyad leads to zero intensity on all layer lines with odd indices (Klug et al. 1958) and so is consistent with the observed layer line spacings, which are all multiples of 1/9.6 nm. On the other hand if the dyad is perpendicular to the fibre axis (Model 3) the structure will belong to line group **sr2** and comprise a pair of oppositely directed helices with a unit height h = 2.4 nm, a unit twist $|t|$ = 90° and a pitch $|P|$ = 9.6 nm. The implications of the latter choice have been studied in considerable detail and many of the conclusions reached are also applicable to Model 2. Subsequent studies suggested that each sheet in Model 3 probably consisted of three complete central strands and two partial outer strands linked together by hairpin bends (Fraser and Parry 2008, 2014), as illustrated in Fig. 6. In Model 2 (Fig. 8), as proposed by Calvaresi et al. (2016), each sheet contains four strands.

8.4 Further Modelling of the Structure of the Filament

On the basis of the conserved nature of the 34-residue β-rich central domain and the out-of-phase eight-residue period in the β- and turn-propensities (Fraser and MacRae 1976; Fraser and Parry 2011a) it is possible to model the central framework of the filaments in some detail. Based on other protein structures, it is to be expected that the internal surface of the β-sheet in the β-keratins would contain predominantly apolar resides and the external side would be more likely to contain charged and hydrophilic residues. A hydropathy study on emu feather and lizard claw (Fraser and Parry 1996), and subsequently on a wide variety of other β-keratin sequences (Fraser and Parry 2011a), did indeed reveal a well-defined oscillation in the hydropathy profile of the proposed β-strands. This is consistent with the hypothesis that the
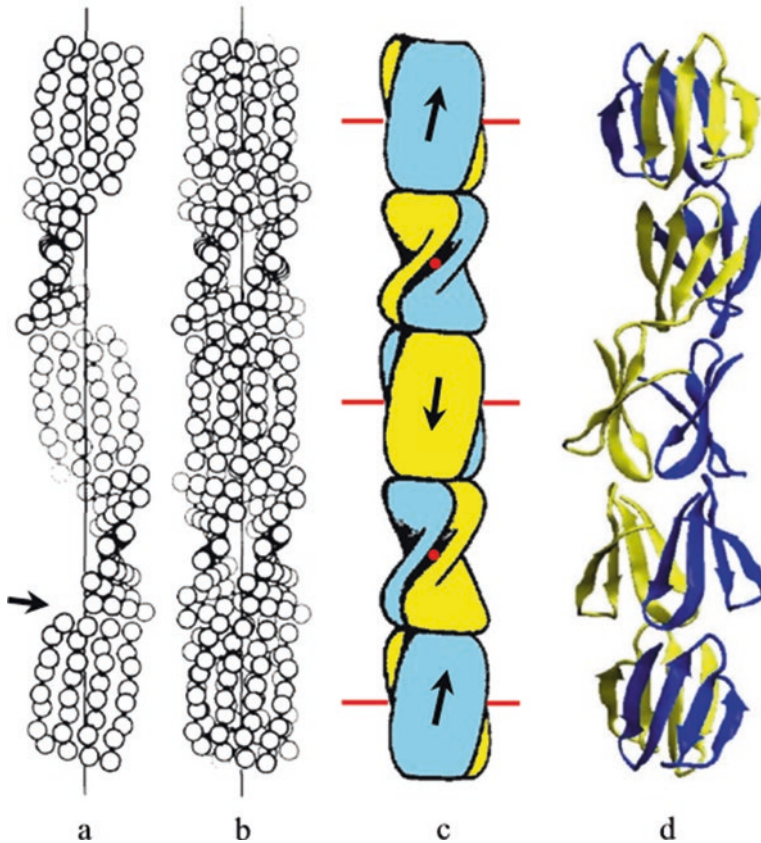
Fig. 8.8 (**a**) The distribution of residues in the string of oppositely-directed sheets in Model 3, derived from infrared and X-ray studies by Fraser et al. (1971). It is important to note that the twist in the sheets is not related to the helical symmetry of the filament (*arrow*), although there must be limitations on possible values if aggregation takes place via interactions between the ends of the sheets. The absence of a relationship between the twist in the sheets and the helical symmetry was confirmed in the refinement of Model 3. (**b**) The filament in this model consists of a pair of oppositely-directed strands of the type shown in (**a**) and the two halves of the so-called β–sandwich (Chothia and Finkelstein 1990; Kister et al. 2002) are related by a perpendicular dyad. The pitch length (*P*) of the two oppositely-directed helices is −9.6 nm i.e. they are left-handed. In this type of sandwich the angle between the chain directions in the two sheets is determined by the extent of their right-handed twist. The observed angle varies between −20° and −50° in protein crystals. (**c**) The relationship between the perpendicular dyads (*red*), and the chain directions (*arrows*) in Model 3. (**d**) The distribution of the β-sandwiches in Model 2 (Calvaresi et al. 2016). The image has been magnified so that the pitch of their basic helix (pitch 2 x 8.5 nm) matches the value derived from the X-ray diffraction pattern (2 x 9.6 nm)

two faces of the β-sheet are chemically distinct. The important role played by the hydropathic/apolar residues in both the assembly of the dimer and in its ultimate stabilization is apparent in the molecular dynamic simulations of Calvaresi et al. (2016). In particular, they showed that the stability of the β-sheet in each chain is

Fig. 8.9 An axial projection of Model 3. (**a**) A single twisted β-sheet, and (**b**) a pair of sheets related by a perpendicular dyad. The side-chains interleave at the apolar interface (Reprinted from Fraser and Parry 2008 with permission from Elsevier)
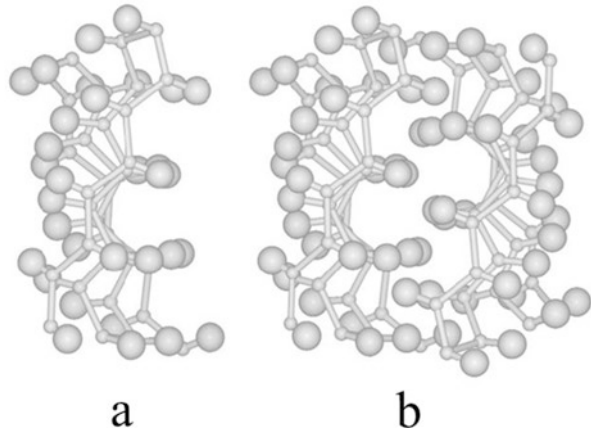
significantly enhanced in an aqueous environment by dimerization. It is important to note, however, that this study gives incorrect values for both the unit height and the pitch in Model 2 and some revisions of the strand length and turn conformations are required.

Modelling of the structure in which the two β-sheets are related by a perpendicular dyad (Model 3) showed that the inner residues in the dimer mesh neatly together in layers oriented perpendicular to the filament axis (Fig. 8.9). This is not so, however, when the sheets are related by a parallel dyad (Model 2).

In studies of Model 3, one observation – a conserved arginine residue in the internal face of the dimer - was unexpected, but does have a rational explanation. Richardson and Richardson (2002) noted that dimeric β-structures in globular proteins, similar to that seen here for β-keratin, often contain an internal arginine residue. Its role is to prevent the exposed edge of the β-sheet bonding to the exposed edge of another β-sheet (Richardson and Richardson 2002). The inner portion of the arginine sidechain is apolar in character and is thus easily compatible with the internal apolar character of the dimer. The charged portion of the arginine residue is at the end of the long sidechain and can therefore be positioned externally to prevent the addition of extra β-strands.

Modelling has also provided some insight into the residues likely to be important in specifying and stabilizing the axial assembly of the dimeric structural units. In Model 3, consecutive dimers differ in rotation about the filament axis by $-90°$ and by 2.4 nm axially. The residues most likely to be involved in both specifying and stabilizing this structure are those central-most bends that connect the β-strands. Four such bends *in toto* are predicted and these occur with n_G values of -18 to -15, -10 to -7, -2 to 1 and 6 to 9. It is therefore of interest that the turn residues absolutely conserved across all chains are the proline residues at $n_G = -9$ and -1, and the glycine at $n_G = 0$. This, allied to the fact that central turns 2 and 3 are very highly conserved across all chains (consensus sequences Q/R-P-P/S-P and L/I-P-G-P, respectively), suggests that these turns, through the provision of strong apolar

interactions, are critical factors in the axial assembly of the filaments (Fraser and Parry 2014). Detailed information on the residues involved in the axial bonding between the β-crystallites in Model 2 have been obtained by Calvaresi et al. (2016) though, as noted earlier, some reconsideration of unit height and pitch in their model is required.

8.5   Packing of Filaments in Sheets in Feather Keratin

Amongst the avian and reptilian epidermal appendages it is thus far only for feather keratin that a level of organization beyond that of the individual filament has been recognized. In this case, small sheets, each containing around 4–10 filaments, have been observed in the electron microscope (Filshie and Rogers 1962). Furthermore, a unique and highly detailed two-dimensional net pattern has been observed in the low-angle X-ray diffraction pattern of feather keratin and this is likewise consistent with a limited sheet-like structure (Bear and Rugo 1951).

The X-ray pattern from the rachis of seagull feather in Fig 8.7a provides strong evidence that the side-to-side packing of filaments is orthogonal with dimensions of 9.5 nm axially and 3.4 nm laterally. This means that specific regions of each filament are lined up equivalently in the direction perpendicular to the filament axes. The electron microscope data likewise indicate that the sheets are relatively flat, but are sufficiently pliant to allow some minor distortion to occur (Filshie and Rogers 1962), possibly as a consequence of the formation of disulphide bonds at the final stage of biosynthesis. In the modelling work of Fraser and Parry (2011b) the orientation of the filaments within the sheets of Model 3 was systematically varied and the diffraction pattern calculated for each model. These were then compared qualitatively against the observed X-ray data. Noting that the individual structural units (dimers) forming one complete pitch length of the filaments are related by a four-fold screw axis and are designated here as 1, 2, 3 and 4, the optimum fit between the models and the low-angle X-ray data occurred when the (local) perpendicular dyad relating structural units 1 and 3 (Fig. 8.8) was inclined at an angle of 17° with respect to the plane of the sheet. (The equivalent angle for units 2 and 4 will be (17–90)° i.e. −73°). In an axial projection it can be seen that these angles result in the β-sheets in one filament lining up with one of the sheets in its immediate neighbour (Fig. 8.10), but the significance of this observation is not immediately clear.

It is of interest that the degree of twist in the individual β-sheets in both Models 2 and 3, was less than that found within most globular proteins, possibly as a result of the large number of tight turns present in the sequences of the filamentous proteins. It is perhaps remarkable that it took some 60 years after the original X-ray observations of Bear and Rugo (1951) before it was possible to interpret their findings in terms of sheets of aligned helical filaments with a specific orientation within the sheets.
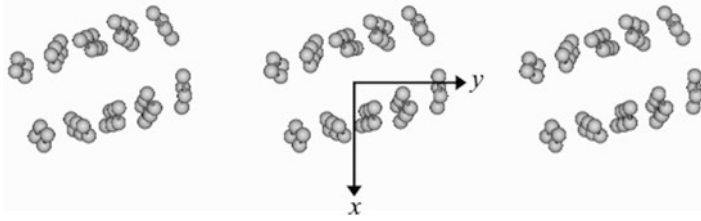
Fig. 8.10 The estimated orientations of adjacent filaments in the sheets responsible for the appearance of row lines in the X-ray diffraction pattern. In this *z*-axis projection of the sheets (Model 3), the orientations of the filaments are such that the β-sandwiches are inclined at an angle of 17° to the plane of the sheet (Reprinted from Fraser and Parry 2011b with permission from Elsevier)

8.6 Physical Properties of β-Keratins and Their Relationship to Sequence Characteristics

There are many examples in biology of filament-matrix composites. The filaments and the matrix each perform a particular role in terms of defining the mechanical and physical properties of the tissue as a whole. In general terms, the filaments provide much of the tensile strength whereas the matrix is associated primarily with its hardness, toughness and pliability (Fraser and Parry 2014). In the human α-keratin filament-matrix composite, for example, multiple protein chains are involved, some of which are found only in the filament (54 chains) and others only in the matrix (89 chains). In contrast, in the hard β-keratins, which form the focus of this review, there is often a single protein family that accounts for the bulk of the appendage, and each member is unique to the particular epidermal appendage in which it occurs. There is little doubt that the mechanical properties of the hard β-keratin appendages must be directly related to the amino acid sequence of its constituent protein. The conserved 34-residue central segment in each chain (and short lengths of sequence about 7 and 11 residues long immediately N- and C-terminal to it respectively) allows filament formation to proceed and this region must therefore be important in specifying tensile strength. The N- and C-terminal domains, however, constitute the matrix and therefore relate directly to the hardness, toughness and pliability of the appendage. The terminal domains differ significantly between appendages and often contain sequence repeats and/or amino acid sequences that are dominated by just a small number of the amino acids (Sect. 8.2). These features have been termed (a) highly charged cysteine-rich, (b) glycine-tyrosine-rich, and (c) glycine-rich. Each would be expected to endow the epidermal appendage with physical properties appropriate to its function *in vivo*. More generally, each appendage will contain a mixture of these segments, which can also differ in length from one chain to another. The hard β-keratins, as a class of structure, are thus of special importance in that they are inherently easier to study than multi-chain composites, and are more

likely to yield insights into the links between sequence, structure and mechanical attributes. However, it must be noted that the relationships are not straightforward, and variables such as the orientation density function of the filaments and the physical form of the appendages will also have a major effect on the physical properties.

8.6.1   Highly Charged Cysteine-Rich Segments

Charged residues are almost invariably located on the surface of a protein and, consequently, lie in positions most easily able to bind water molecules. It follows that these regions are likely to be more heavily hydrated than those regions lacking significant numbers of charged residues. Taylor et al. (2004) have studied the effect of hydration on the tensile properties of avian keratins and concluded that an increase in water content will soften the tissue and render it more pliable. The role of the high number of cysteine residues present will result in additional disulphide bridges being formed. These will ensure that the insolubility of the tissue and its resistance to its environment are not compromised.

8.6.2   Glycine-Tyrosine-Rich Segments

The glycine-tyrosine-rich segments, by analogy with those characterized in the hard α-keratins, are also referred to as high glycine-tyrosine segments (HGT). Glycine and tyrosine both display particular structural features of interest. Glycine is the smallest amino acid and is apolar in character. As such it normally occurs in the interior of a protein structure. Having such a small sidechain (a single hydrogen atom) glycine has much more conformational freedom than any other amino acid and, as such, a glycine-rich region may readily adopt a variety of energetically-equivalent conformations stabilized by apolar interactions (Fraser and Parry 2014). Tyrosine residues, in contrast, are very much larger than glycines and contain an aromatic ring and a reactive hydroxyl group. Like glycine, however, they are more usually found in protein interiors than on their surfaces. It is very probable that tyrosine residues interact strongly with one another either via ring stacking (McGaughey et al. 1998) or through the formation of hydrogen bonds between centre of the aromatic ring and the tyrosine's hydroxyl group (Levitt and Perutz 1988). Thus, having a region in a β-keratin containing highly interactive tyrosine residues linked by conformationally-free glycine residues is expected to result in the material being strengthened considerably whilst simultaneously having a degree of pliability.

8.6.3   Glycine-Rich Segments

These were first defined for the epidermal α-keratins, where segments in both the N- and C-terminal domains of the intermediate filament molecules (known as the V1 and V2 subdomains respectively) were shown to be very rich in glycine (and often serine) residues. Steinert et al. (1991) proposed that these subdomains would form glycine loops anchored on interactions between large apolar residues that were also present in these sequences. The concept was thus established of loops of variable conformation that were easily extended. However, because of their essentially apolar character, it was believed that on energetic grounds the loops would generally prefer to adopt a more compact form, thereby shielding as many of them as possible from the aqueous environment. These glycine-rich segments may thereby endow the tissue with a degree of elasticity.

8.7   Summary

Many of the key structural features of the proteins that form the epidermal appendages of birds and reptiles are now well established. From analyses of the several hundred amino acid sequences now available it has been shown that all chains contain a highly conserved region of 34-residues and that this adopts a twisted antiparallel β-sheet conformation (believed to be right-handed) with one face largely composed of apolar residues. Three possibilities exist for the symmetry of the pair of sheets forming the β-sandwich. In Model 1, the two sheets have different conformations, resulting in a filament belonging to line group **s**, in Model 2, they have identical conformations and are related by a dyad parallel to the fiber axis leading to a structure belonging to line group **sr** (*N*=2), and in Model 3, the two sheets have identical conformations but are related by a dyad axis perpendicular to the fibre axis leading to a structure belonging to line group **sr2**. Progress in the determination of the three-dimensional structure of feather keratin is summarized in Fig. 8.11, which illustrates the relationship between the details seen in electron microscopy studies and the current model for the three-dimensional structure.

   A word of caution is necessary about the model-building method, as used in developing a model for avian and reptilian keratins, as opposed to the crystallization method, which gives an electron density map of the *actual* structure. In the former case, the function of the model is to highlight discrepancies between the observed and predicted properties and to form a basis for further development. In practice, any refinement of a model is unlikely to be a unique solution and, with increasing complexity, the number of parameters needed to define the model multiplies rapidly. The problems involved in searching for the global minimum of a merit function in *n*-dimensional space as *n* increases are well documented and it is necessary to view all "progress" in model refinement *cum grano salis*.
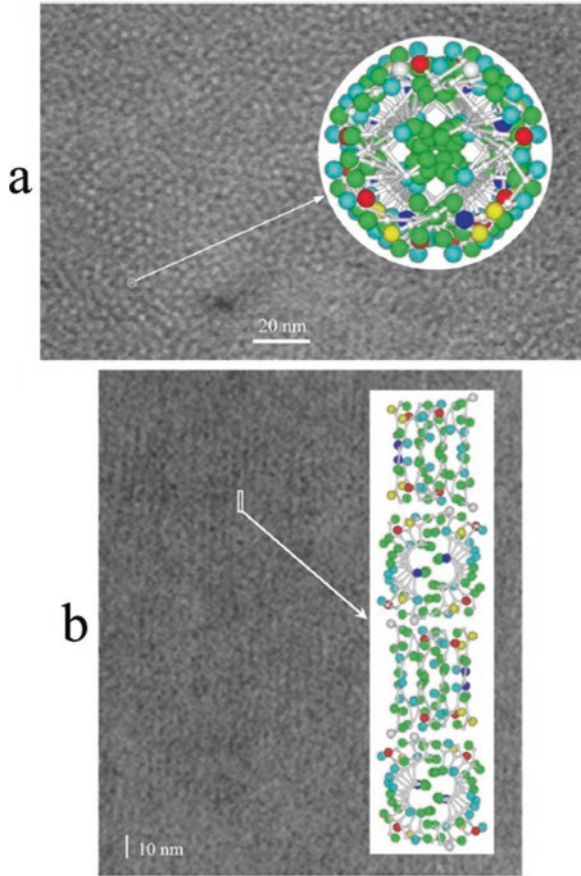
Fig. 8.11 The relationship between the detail seen in electron microscope studies of feather keratin and Model 3 for the three-dimensional structure. (**a**) Transmission electron micrograph of a cross-section of feather (Filshie and Rogers 1962) with an axial projection of the proposed structure superposed. The diameter of the superposed circle is 2.4 nm and the sidechain colours are *green* for hydrophobic, *pale blue* for hydrogen-bonding, *red* for acidic, *dark blue* for basic, *yellow* for half-cystine and *grey* for glycine. It will be seen that the core of the filament is composed of hydrophobic residues whilst the charged and half-cystine residues are found towards the exterior of the filament. This would also be the case in Model 2. The diameter of the filament framework matches that of the unstained area in the micrograph. A similar match was found with micrographs obtained from chicken leg scale (Stewart 1977). (**b**) Micrograph of a longitudinal section of seagull feather (Filshie and Rogers 1962) showing clearly resolved filaments. A longitudinal structure is apparent, but it is difficult to assess the periodicity. Inset: longitudinal view of the model structure showing that the dimers have the correct axial length to join up to give a continuous filament (Reprinted from Fraser and Parry 2008 with permission from Elsevier)

In the case of feather keratin 4–10 of these 3.4 nm diameter β-keratin filaments form a sheet in which the specific orientation of the constituent filaments has been determined by modelling against the observed low-angle X-ray diffraction pattern.

The N- and C-terminal domains of the single chain species present contain sub-domains of various lengths that have special sequence/composition characteristics such a glycine-rich, glycine-tyrosine-rich and charged residue-cysteine residue-rich. The Squamates and the Archosaurs each contain common features in terms of the presence of specific subdomains, and in many cases the relationship between the nature of the subdomains and the mechanical and physical properties of the particular epidermal appendage has been recognized.

The complex structure of feather rachis is assembled in the follicle from a single molecule containing, in the case of the mallard duck, only 95 residues. Early studies showed that the disulphide bonds in seagull rachis can be reduced and substituted to give a soluble protein that, under suitable conditions, can be reassembled to give filaments with an X-ray diffraction pattern having a strong 2.4 nm meridional reflection (Filshie et al. 1964). This points to a possible way to further progress, with the exciting possibility of modifying soluble derivatives or synthesizing 'tailored' molecules that can be crystallized and used to gain information on the electron density distribution and the actual β-conformation in the strands and loops in the central domain.

References

Alibardi L, Toni M (2007) Beta-keratins of reptilian scales share a central amino acid sequence termed core-box. Res J Biol Sci 2:329–339

Alibardi L, Dalla Valle L, Toffolo V, Toni M (2006) Scale keratin in lizard epidermis reveals amino acid regions homologous with avian and mammalian proteins. Anat Rec 288A:734–752

Alibardi L, Dalla Valle L, Toni M (2009) Cell biology of adhesive setae in gecko lizards. Zoology 112:403–424

Alexander NJ (1970) Comparison of alpha and beta keratins in reptiles. Z. Zellforsch. Mikrosk Anat 110:153–165

Astbury WT, Marwick TC (1932) X-ray interpretation of the molecular structure of feather keratin. Nature (London) 130:309–310

Bear RS, Rugo HJ (1951) The results of X-ray diffraction studies on keratin fibers. Ann N Y Acad Sci 53:627–648

Calvaresi M, Eckhart L, Alibardi L (2016) The molecular organization of the beta-sheet region in corneous beta-proteins (beta-keratins) of sauropsids explains its stability and polymerization into filaments. J Struct Biol 194:282–291

Chothia C, Finkelstein AV (1990) The classification and origins of protein folding patterns. Annu Rev Biochem 57:1007–1039

Chou K (2000) Prediction of tight turns and their types in proteins. Anal Biochem 286:1–16

Dalla Valle L, Toffolo V, Belvedere P, Alibardi L (2005) Isolation of a mRNA encoding a glycine-proline-rich β-keratin expressed in the regenerating epidermis of lizard. Dev Dyn 234:934–947

Dalla Valle L, Nardi A, Belvedere P, Toni M, Alibardi L (2007a) β-Keratins of differentiating epidermis of snake comprise glycine-proline-serine-rich proteins with an avian-like gene organization. Dev. Dynamics 236:1939–1953

Dalla Valle L, Nardi A, Toffolo V, Niero C, Toni M, Alibardi L (2007b) Cloning and characterization of scale β-keratins in the differentiating epidermis of geckoes show they are glycine-proline-serine-rich proteins with a central motif homologous to avian keratins. Dev Dyn 236:374–388

Dalla Valle L, Nardi A, Toni M, Alibardi L (2009a) Beta-keratins of turtle shell comprise small glycine-proline-tyrosine rich proteins similar to those of crocodilians and birds. J Anat 214:284–300

Dalla Valle L, Nardi A, Gelmi C, Toni M, Emera D, Alibardi L (2009b) β-keratins of the crocodilian epidermis: composition, structure, and phylogenetic relationships. J Exp Zool (Mol Dev Evol) 312B:42–57

Dalla Valle L, Nardi GB, Bonazza G, Zuccal C, Emera D, Alibardi L (2010) Forty keratin-associated β-proteins (β-proteins) form the hard layers of scales, claws, adhesive pads in the green anole lizard, *Anolis carolinensis*. J Exp Zool 314B:11–32

Filshie BK, Rogers GE (1962) An electron microscope study of the fine structure of feather keratin. J Cell Biol 13:1–12

Filshie BK, Rogers GE, Fraser RDB, MacRae TP (1964) X-ray diffraction and electron-microscope observations on soluble derivatives of feather keratin. Biochem J 92:18–19

Fraser RDB, MacRae TP (1959) Molecular organization in feather keratin. J Mol Biol 1:387–397

Fraser RDB, MacRae TP (1963) Structural organization in feather keratin. J Mol Biol 7:272–280

Fraser RDB, MacRae TP (1976) The molecular structure of feather keratin. In: Proceedings of the 16th international ornithological congress, Canberra 443–451

Fraser RDB, Parry DAD (1996) The molecular structure of reptilian keratin. Int J Biol Macromol 19:207–211

Fraser RDB, Parry DAD (2008) Molecular packing in the feather keratin filament. J Struct Biol 162:1–13

Fraser RDB, Parry DAD (2011a) The structural basis of the filament-matrix texture in the avian/reptilian group of hard β-keratins. J Struct Biol 173:391–405

Fraser RDB, Parry DAD (2011b) The structural basis of the two-dimensional net pattern observed in the X-ray diffraction pattern of avian keratin. J Struct Biol 176:340–349

Fraser RDB, Parry DAD (2014) Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties. J Struct Biol 188:213–224

Fraser RDB, Suzuki E (1965) Polypeptide chain conformation in feather keratin. J Mol Biol 14:279–282

Fraser RDB, MacRae TP, Parry DAD, Suzuki E (1971) The structure of feather keratin. Polymer 12:35–56

Greenwold MJ, Sawyer RH (2010) Genomic organization and molecular phylogenies of the beta (β) keratin multigene family in the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*): implications for feather evolution. BMC Evol Biol 10:148–157

Gregg K, Wilton SD, Parry DAD, Rogers GE (1984) A comparison of genomic coding sequences for feather and scale keratins: structural and evolutionary implications. EMBO J 3:175–178

Hallahan DL, Keiper-Hrynko NM, Shang TQ, Ganzke TS, Toni M, Dalla Valle L, Alibardi L (2009) Analysis of gene expression in gecko digital adhesive pads indicates significant production of cysteine- and glycine-rich beta keratins. J Exp Zool 312B:58–73

Inglis AS, Gillespie MJ, Roxborough CM, Whitaker LA, Casagranda F (1987) In: L'Italien JL (ed) Proteins, structure and function. Plenum, New York, pp 757–764

Kister AE, Finkelstein AV, Gelfand IM (2002) Common features in structures and sequences of sandwich-like proteins. Proc Natl Acad Sci U S A 99:14137–14141

Klug A, Crick FHC, Wyckoff HW (1958) Diffraction by helical structures. Acta Cryst 11:199–213

Landmann L (1979) Keratin formation and barrier mechanisms in the epidermis of *Natrix natrix* (Reptilia, serpentes): an ultrastructural study. J Morphol 162:93–126

Levitt M, Perutz MF (1988) Aromatic rings act as hydrogen-bond acceptors. J Mol Biol 201:751–754

Maderson PFA, Flaxman BA, Roth SI, Szabo G (1972) Ultrastructural contribution to the identification of cell types in the lizard epidermal generation. J Morphol 136:191–209

McGaughey GB, Gagne M, Rappe AK (1998) π-Stacking interactions alive and well in proteins. J. Biol Chem 273:15458–15463

O'Donnell IJ (1973) The complete amino acid sequence of a feather keratin from emu (Dromaius novae-hollandiae). Aust J Biol Sci 26:415–437

O'Donnell IJ, Inglis AS (1974) Amino acid sequence of a feather keratin from silver gull (*Larus novae-hollandiae*) and comparison with one from emu (*Dromaius novae-hollandiae*). Aust J Biol Sci 27:369–382

Rest JS, Ast JC, Austin CC, Wadell PJ, Tibbetts EA, Hay JM, Mindell DP (2003) Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. Mol Phylogenet Evol 29:289–297

Richardson JS, Richardson DC (2002) Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci U S A 99:2754–2760

Rudall KM (1947) X-ray studies of the distribution of protein chain types in the vertebrate epidermis. Biochim Biophys Acta 1:549–562

Sawyer RH, Glenn T, French JO, Mays B, Shames RB, Barnes GL, Rhodes W, Ishikawa Y (2000) The expression of beta (β) keratins in the epidermal appendages of reptiles and birds. Am Zool 40:530–539

Schorr R, Krimm S (1961) Studies on the structure of feather keratin I. X-ray diffraction studies and other experimental data. Biophys J 1:467–487

Steinert PM, Mack JW, Korge BP, Gan S-Q, Haynes SR, Steven AC (1991) Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. Int J Biol Macromol 13:130–139

Stewart M (1977) The structure of chicken scale keratin. J Ultrastruct Res 60:27–33

Suzuki E (1973) Localization of beta-conformation in feather keratin. Aust J Biol Sci 26:435–437

Taylor AM, Bonser RHC, Farrent JW (2004) The influence of hydration on the tensile and compressive properties of avian keratinous tissues. J Mater Sci 39:939–942

Chapter 9
Tropomyosin Structure, Function, and Interactions: A Dynamic Regulator

Sarah E. Hitchcock-DeGregori and Bipasha Barua

Contents

Abstract Tropomyosin is the archetypal-coiled coil, yet studies of its structure and function have proven it to be a dynamic regulator of actin filament function in muscle and non-muscle cells. Here we review aspects of its structure that deviate from canonical leucine zipper coiled coils that allow tropomyosin to bind to actin, regulate myosin, and interact directly and indirectly with actin-binding proteins. Four genes encode tropomyosins in vertebrates, with additional diversity that results from alternate promoters and alternatively spliced exons. At the same time that periodic motifs for binding actin and regulating myosin are conserved, isoform-specific domains allow for specific interaction with myosins and actin filament regulatory

S.E. Hitchcock-DeGregori (✉) • B. Barua
Department of Pathology and Laboratory Medicine, Robert Wood Johnson Medical School,
Rutgers University, 675 Hoes Lane West, Piscataway, NJ 08854, USA
e-mail: hitchcoc@rwjms.rutgers.edu

proteins, including troponin. Tropomyosin can be viewed as a universal regulator of the actin cytoskeleton that specifies actin filaments for cellular and intracellular functions.

Keywords Tropomyosin • Actin • Myosin • Coiled coil • Actin cytoskeleton • Myosin regulation • Actin filament dynamics

9.1   Introduction

When Kenneth Bailey reported his discovery of tropomyosin in 1946 as "an unusual protein of fibrous character … isolated from the fibrillar fraction of skeletal and cardiac muscle," he had no idea how unusual tropomyosin is (Bailey 1946). Far from being an "inert" structural protein, as fibrous proteins have often been imagined to be, tropomyosin is a two-chained α-helical coiled-coil actin binding protein that has active roles in the regulation of actin filament dynamics and function in eukaryotic cells. Research over decades has given us insight into how tropomyosin achieves this, and has posed new questions to answer.

Bailey classified tropomyosin as a fibrous protein, based on the positive flow birefringence and high relative viscosity of the purified protein (Bailey et al. 1948). Wide-angle X-ray patterns revealed the α-type pattern typical of the *k-m-e-f* (keratin-myosin-epidermis-fibrinogen) group of proteins that we now know are α-helical coiled-coil proteins (Astbury et al. 1948). The extensive characterization of the protein was remarkable for the time. It included the molecular weight, amino acid composition, revealing its highly charged nature and the unusual ratio between polar and non-polar residues, and allowed the (correct) suggestion that tropomyosin has two chains (Bailey 1948, 1951; Bailey et al. 1948; Tsao et al. 1951). These early physical and chemical characterizations, together with his knowledge of fiber diffraction, gave Francis Crick what he needed to postulate his model of the coiled coil, not long after Linus Pauling deduced the structure of the α-helix (Pauling et al. 1951). Two right-handed amphipathic α-helices wind around each other to form a left-handed supercoil (see Chap. 1), with a hydrophobic strip of amino acids packed "knobs-into-holes" between two polypeptide chains in which hydrophobic side-chains of one chain ("knobs") pack against depressions ("holes") on the surface of the α-helix of the other chain (Fig. 9.1) (Crick 1953).

Many years later, the amino acid sequence of striated muscle tropomyosin was determined chemically from native protein, before the invention of recombinant DNA and DNA sequencing (Sodek et al. 1972, 1978). As the first fibrous protein to be fully sequenced, knowledge of the sequence moved the field forward by confirming Crick's 1953 proposal. It revealed that tropomyosin is a coiled coil along its length: the repeat of hydrophobic residues at the *a* and *d* interface positions of the heptad is uninterrupted (McLachlan and Stewart 1975; McLachlan et al. 1975; Parry 1975). The X-ray structure of the GCN4 "leucine zipper" with the interlocking knobs-into-holes packing of the hydrophobic side chains in the interface between the two α-helices proved Crick's model (O'Shea et al. 1991). Analysis of the
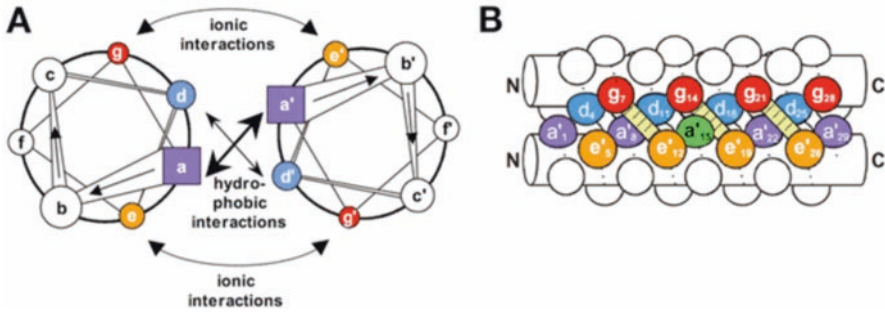
Fig. 9.1 Model of a two-chained, parallel coiled coil such as that in tropomyosin. (**A**) Helical wheel representation viewing the helix axis from the C terminus. (**B**) Side view in which the helical backbones are represented by cylinders. The positions are labeled *a* to *g* and *a'* to *g'* on the two helices. The interacting residues that stabilize the coiled coil are colored or shaded: *a* and *d* form a hydrophobic interface. *e* and *g* pack against the hydrophobic core and can form inter-helical salt bridges if oppositely charged; the *g* residue can form a salt bridge between the oppositely charged *e'* residue in the next heptad of the other helix (Modified with permission from Mason and Arndt (2004))

sequence also gave insights into functional design. The amino acid sequence contains periodic repeats with periodicities attributed to gene duplication as well as sevenfold and 14-fold periodicities of surface residues predicted to be quasi-equivalent actin binding sites (McLachlan and Stewart 1976; McLachlan et al. 1975; Parry 1975; Parry and Squire 1973; Stewart and McLachlan 1975).

Bailey named the protein tropomyosin "not so much to describe its [unknown] function … but to imply that in all its properties it is a prototype of myosin" (Bailey 1948). Tropomyosin is a prototype, if not the way Bailey envisioned. Tropomyosin is the archetypal-coiled coil as it relates to protein structure and design, versus being a "primitive" form. Hodges and his colleagues used tropomyosin model peptides to test and define the fundamental parameters for coiled-coil assembly (e.g. Hodges et al. 1981; Tripet et al. 2000; Wagschal et al. 1999a, b). The coiled coil is now recognized to be a wide-spread structural motif involved in oligomerization of many protein complexes (Parry et al. 2008). Here we aim to show you that tropomyosin is not a strict Crick coiled coil.

In the context of this volume on fibrous proteins, we focus on aspects of tropomyosin structure and function of particular interest to us, that we believe show why it continues to be fascinating to those of us who have tried to unlock its secrets, only to reveal more unanswered questions, as is often the case in science. We review the structure of tropomyosin and discuss features that depart from those of canonical coiled coils that are postulated to enable specific functions, in particular for tropomyosin to bind along the length of the helical actin filament. We discuss the evolutionarily conserved design of actin binding sites and myosin regulatory sites in relation to recent molecular models. Critical functions of tropomyosin include the allosteric, cooperative regulation of the actin filament with myosin, and the striated muscle Ca^{2+} sensor – troponin – that binds both actin and tropomyosin. We focus on the universality and specificity of myosin regulation, leaving thin filament regulation by

troponin to (Paul et al. 2017). Finally, we will illustrate that knowledge of the multiplicity of tropomyosin genes and the isoforms they encode has opened our eyes to how tropomyosin serves as a universal regulator of the actin cytoskeleton and its dynamics in diverse cellular functions by differentiating actin tracks within cells (Gunning et al. 2015; Khaitlina 2015; Wang and Coluccio 2010). For additional perspectives beyond what we are able to cover in this chapter we refer you to *Tropomyosin* (Gunning 2008) and the February, 2014 issue of *The Journal of Muscle Research and Cell Motility* that is devoted to tropomyosin, and other reviews cited therein. Finally, mutations in tropomyosins are the cause of cardiac and skeletal myopathies (Kee and Hardeman 2008; Redwood and Robinson 2013; Tardiff 2011), and a hypersensitivity response to tropomyosin a the cause of shellfish and dust allergies (Wong et al. 2016).

9.2 The Structure of Tropomyosin

Until the development of modern analytical methods for proteins, crystallization of newly discovered proteins was expected as a standard of purity. In that light, Bailey reported thin, fragile tropomyosin crystals that he noted were "enormously" hydrated, ~90 % water, more than twice that of other protein crystals at that time (Bailey 1946, 1948). Indeed, the high water content has continued to plague crystallographic studies of tropomyosin. Despite being the first sarcomeric protein to be crystallized, the highest resolution structure of a full-length tropomyosin to date is 7 Å (Fig. 9.2A) (Whitby and Phillips 2000). Nevertheless, the lower resolution

Fig. 9.2 (continued) is oriented to highlight the bends in the crystal structure of the full-length tropomyosin molecule. (**B**) Models of segments of the crystal structure of the first 80 residues of tropomyosin (PDB ID 1IC2) show the influence of the interface residues on packing, axial stagger, and interhelical distance. The chains with an Ala cluster (*left*) are staggered, the interhelical distance is closer by ~2 Å, and the interface is poorly packed compared to the canonical coiled coil interface (*right*); superimposed in the middle (Modified from Brown et al. (2001)). (**C**) While regions of tropomyosin with Ala clusters are close-packed, isolated core alanines create holes in the coiled coil. The sizes of the sequential interface side chains influences the packing of the interface. The residue numbers refer to regions in the PDB ID 2B9C structure (Modified with permission from Brown et al. (2005)). (**D**) Model of the structure of the C-terminal third of tropomyosin, stabilized at the N terminus (2D3E). The side-chains of the destabilizing interface residues Tyr214-Glu218-Tyr221 are illustrated. The location of this interface at the site of a bend in the coiled coil is illustrated on the left; the increased interhelical distance is illustrated on the right in a model rotated by 90°. (Modified with permission from Nitanai et al. (2007)). (**E**) Parameters of the coiled coil included in a local flexibility analysis of tropomyosin using molecular dynamics simulations. (**F**) Results of local flexibility analysis of tropomyosin compared with x-ray structures of tropomyosin. The MD simulations starting with a model coiled coil (*blue circles*) reproduce the fluctuations in interhelical distances seen in crystal structures (other colors). Open circles represent the initial model (Li et al. 2011). The *magenta vertical bars* on the abscissa are the sites of interface alanines; the *grey vertical bars* are the sites of Ala clusters where the chains are closer together. (**G**) The MD data from F. are reproduced here with the average (*blue*) and SD (standard deviation) (*red*). The greatest SD is at the site of a bend where there is an interface Glu (Fig. 9.2D): (E, F, G, modified from Zheng et al. (2013))
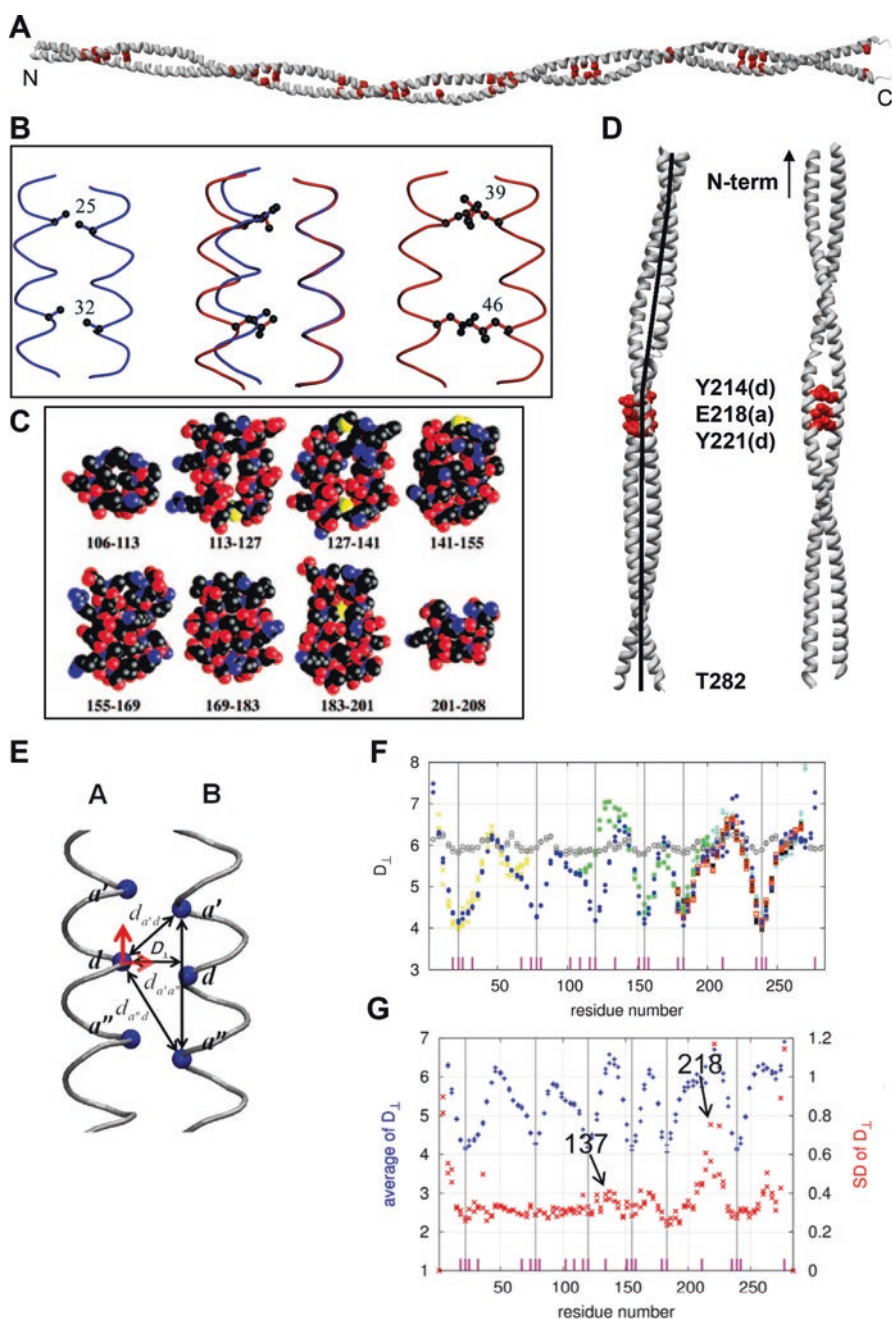
Fig. 9.2 The variable and dynamic tropomyosin coiled coil. Panels (**A–D**) illustrate how non-canonical interface residues of the coiled coil influence packing and interhelical distance. (**A**) The side-chains of the interface (*a, d* positions) alanines shown on a ribbon model of the 7 Å structure (PDB ID 1C1G) (Whitby and Phillips 2000) illustrate their quasi-periodic distribution. The model

structures and the study of paracrystals in the electron microscope defined certain basic features of muscle tropomyosin: the molecular length, the approximate length of the N-to-C-terminal junction, the parallel approximately in-register alignment of two chains, the flexibility and presence of bends, and the variable pitch length of the supercoil (Brown and Cohen 2005; Caspar et al. 1969; Perry 2001; Phillips et al. 1979). Concurrent biophysical studies indicated that tropomyosin is not a canonical "Crick" coiled coil along its entire length; it is a dynamic protein with domains that fold and unfold independently of one another, reflecting regions of greater and lesser stability (Brown and Cohen 2005; El-Mezgueldi 2014; Hitchcock-DeGregori 2008; Perry 2001).

We focus here on more recent NMR and X-ray studies of fragments and peptides of muscle tropomyosins that have yielded atomic resolution structures (Table 9.1). Together these structures span most of the length of tropomyosin. Much can be deduced about the structure from the heptapeptide repeat in the sequence, including distortions resulting from interruptions in the pattern of core hydrophobic residues and stabilizing inter-helical salt bridges. The structures, coupled with experimental and computational investigations and model building have provided insights into the intricate design of tropomyosin for its functions.

Even so, the variations from a canonical coiled coil revealed in the atomic resolution NMR and X-ray structures were stunning and generally unanticipated (Hitchcock-DeGregori 2008). There are variations in inter-helical distance; staggers between the helices, bends, and holes or spaces in the interface that in some cases contain water molecules that interrupt canonical regions of the coiled coil. We will focus here on the variations in coiled coil structure, how they make tropomyosin semi-flexible and provide insights into function. The bends, staggers, and variations in the length of the supercoil illustrate the dynamic nature of the tropomyosin coiled coil that allows it to bind to the helical actin filament and change positions on the filament in response to myosin binding, Ca^{2+} binding to troponin, or binding of another actin filament regulator.

9.2.1   Alanine Clusters

The tropomyosin sequence has an unusually high number of interface alanines for a coiled coil, residues that do not enable canonical Crick-like knobs-into-holes packing at the hydrophobic interface. McLachlan and Stewart noticed that regions with a canonical interface (Leu, Ile) were interrupted with Ala-rich regions (Fig. 9.2A) (McLachlan and Stewart 1976). The consequence became clear in the structure of the first 80 residues of striated muscle tropomyosin (Brown et al. 2001). In regions with predominantly Ala in the core, so-called "Ala clusters", the chains are axially out-of-register by about 1 Å and are about 2 Å closer together than in canonical regions. Hence the interface remains relatively well close-packed. The resulting stagger causes a directionally specific, near in-plane bend in the coiled coil (Fig. 9.2B) (Brown 2010; Brown et al. 2001). Such geometric-specific bends are

Table 9.1 Tropomyosin structures

Description	PDB ID	Region	Modifications	References
Vertebrate muscle tropomyosins				
Rat Tpm1.1, 15 Å	2TMA, X-ray	1–284	Native	Phillips et al. (1986)
Pig Tpm1.1, 7 Å	1C1G, X-ray	1–284	Native	Whitby and Phillips (2000)
N terminus, Tpm1.1	TM1Z, NMR	1–14	N-acetylated, 18 residue Leu zipper on C terminus	Greenfield et al. (1998)
N terminus, Tpm1.1	1IC2, X-ray	1–80	N terminus unacetylated, Residue 81 is Cys, S-S	Brown et al. (2001)
N terminus, Tpm1.2	3U1A, X-ray	1–81	AGH at N terminus	Rao et al. (2012)
	3U1C, X-ray	1–98	AGH at N terminus, antiparallel dimer	
N terminus, Tpm2.1	3U59, X-ray	1–98	GAS at N terminus, four chains, overlap	Rao et al. (2012)
Mid-region, Tpm1.1	2B9C, X-ray	89–208	Fused to GCN4 Leu zipper at C terminus	Brown et al. (2005)
C terminus, Tpm1.1	2DE3, X-ray	176–284	GCN4 at N terminus	Minakata et al. (2008)
	2EFR, 2EFS	176–273	GCN4 at both ends	Nitanai et al. (2007)
C terminus, Tpm1.1	1KGL, X-ray	254–284	Fused to GCN4 Leu zipper at N terminus	Li et al. (2002)
C terminus, Tpm1.1	1MV4, NMR	252–284	GCG at N terminus, S-S	Greenfield et al. (2003)
Overlap, Tpm1.1	2G9J, NMR	1–14, 254–284	N terminus GCN4, N-Gly; C terminus GCG, S-S	Greenfield et al. (2006)
Overlap, Tpm2.1	3MUD, X-ray	1–29, 256–284	N terminus GAS 1-29 EB1; C terminus Gp7 256-284	Frye et al. (2010)
	3MTU, X-ray	1–29, 248–284	N terminus GAS 1-29 EB1; C terminus XRCC4 248-284	
Overlap, Tpm1.1	2Z5I, X-ray	1–24, 254–284	N terminus 1-24, GCN4, GAS; C terminus GCN4 254-284	Murakami et al. (2008)
Vertebrate non-muscle tropomyosins				
N terminus, 1b, TPM1	1IHQ, NMR	1–19	N-terminus Gly, GCN4 zipper at C terminus	Greenfield et al. (2001)
N-terminus, 1b, TPM1	3AZD, X-ray	1–19	N-terminus Gly, GCN4 zipper at C terminus	Meshcheryakov et al. (2011)
Overlap, 1b, 9d, TPM1	2K8X, NMR	1–19, 252–284	N-terminus Gly, GCN4 at C-terminus, 9d unlabeled	Greenfield et al. (2009)

present in X-ray structures of other regions of tropomyosin that contain close-packed Ala clusters (Fig. 9.2C) (Brown et al. 2005; Minakata et al. 2008). Ala clusters are distributed quasi-periodically in the sequence in a conserved, and functionally important, relationship to the proposed seven quasi-equivalent actin binding sites (see Sect. 9.3.1). The bends contribute to the semi-flexibility and variable supercoil that enable the tropomyosin to wind around the helical actin filament (Singh and Hitchcock-DeGregori 2003; Singh and Hitchcock-DeGregori 2006). In addition, the altered packing and stagger at the Ala clusters break the symmetry of those tropomyosin coiled coils whose parallel polypeptide chains are identical in sequence.

9.2.2   Bends and Holes in Core Packing

There are two locations in striated muscle tropomyosin where acidic residues at interface positions can cause bends in the coiled coil: Asp137 in a *d* position and Glu218 in an *a* position (Fig. 9.2D) (Brown et al. 2005; Minakata et al. 2008; Nitanai et al. 2007). These two residues are among the most highly conserved in animal tropomyosins (Barua et al. 2011). The bends may also have a broken core, with a hole, that in some cases contains bound water (Fig. 9.2C) (Brown et al. 2005; Minakata et al. 2008). These and other regions of tropomyosin with non-canonical cores, or a core Ala surrounded by large core resides, result in holes (Fig. 9.2C) and are in regions with high B-factors, indicative of dynamic flexibility or local instability, and out-of-plane bends (Brown 2010; Brown et al. 2005; Minakata et al. 2008). Mutation of Asp137 to a canonical Leu has long-range effects within tropomyosin and influences actin filament stiffness (Matyushenko et al. 2014; Sumida et al. 2008). The effects of particular combinations of core residues on the structure correlate well with the effects on stability of model coiled coil peptides (Kwok and Hodges 2003; Lu and Hodges 2004). Molecular dynamics simulations of a tropomyosin molecule starting with a canonical coiled coil conformation have replicated the variations in inter-chain distance and bends observed in crystal structures (Fig. 9.2E-G), reassuring that the deviations from canonical coiled-coil are not solely artifacts of crystal packing or modifications of the tropomyosin fragments to stabilize the coiled coil for structural studies.

9.2.3   The Overlap Complex

Tropomyosin polymerizes head-to-tail along both sides of the actin filament to form a continuous parallel coiled-coil cable in which the N terminus and C terminus overlap by 11–14 residues. To understand how the actin binding sites on sequential tropomyosin molecules along the filament present to the actin subunits it was

important to know the structure of the overlap complex. The ends of tropomyosin are essential functional domains and the source of isoform specificity (see Sect. 9.5).

How does the coiled coil N terminus bind to a C terminus that is flexible? The atomic resolution structures of complexes of N-terminal and C-terminal model peptides were unanticipated, and are key to understanding how sequential tropomyosin molecules on the filament may bind to actin. The C-terminal helices are parallel in solution and splay apart, as in the x-ray structures of the C terminus, to accommodate the N-terminal coiled coil (Greenfield et al. 2003, 2006; Li et al. 2002; Minakata et al. 2008). The N and C termini overlap by about 11 residues, stabilized by hydrophobic and ionic interactions, to form an unusual four-helix bundle. A significant feature is that the plane of the N-terminal coiled coil is rotated 90° relative to the plane of the C terminus (Fig. 9.3A). The junction is flexible, as illustrated by the ensemble of NMR structures, with a directional bias that is presumably crucial for the tropomyosin supercoil to conform to and move on the helical actin filament. X-ray structures of the overlap complex of smooth muscle tropomyosin confirmed the solution structure, as did molecular dynamics simulations of full-length tropomyosin (Frye et al. 2010; Li et al. 2014). All three differ in the degree of curvature, illustrative of the semi-flexibility of the junction.

9.2.4 Tropomyosin Dynamics and Transmission of Structural Information

Some of the variations from canonical coiled-coil structure might be predicted from models and molecular dynamics simulations. However, the complexity of the dynamics and long-range transmission of stability information in tropomyosin are more difficult to understand. The flexibility was evident from early studies of polymorphic crystal forms (Caspar et al. 1969; Cohen et al. 1972). The observation of molecular motions perpendicular to the coiled-coil axis in crystals allowed the suggestion that aperiodic features may allow flexibility and motions that are important for function (Phillips et al. 1980, 1986). The structural studies were corroborated by solution studies that indicated regions of local instability and multiple unfolding domains that can now be at least partially explained in structural terms (Graceffa and Lehrer 1980; Ishii et al. 1992; Potekhin and Privalov 1982).

An important question that remains is, how is structural and stability information cooperatively transmitted through the coiled coil? It has long been known that protein modifications, such as the introduction of a fluorescence probe or a disulfide bond, for example, can have long-range effects (>100 Å) within the molecule that are manifested by local unfolding, protease-sensitivity or regulatory function, e.g. (Graceffa and Lehrer 1980; Singh and Hitchcock-DeGregori 2009; Ueno 1984). However, how conformational changes are dynamically transmitted within the coiled coil is unknown. Based on an analysis of staggers in structures of coiled coils, Brown suggests that structural information can be directionally transmitted for >40
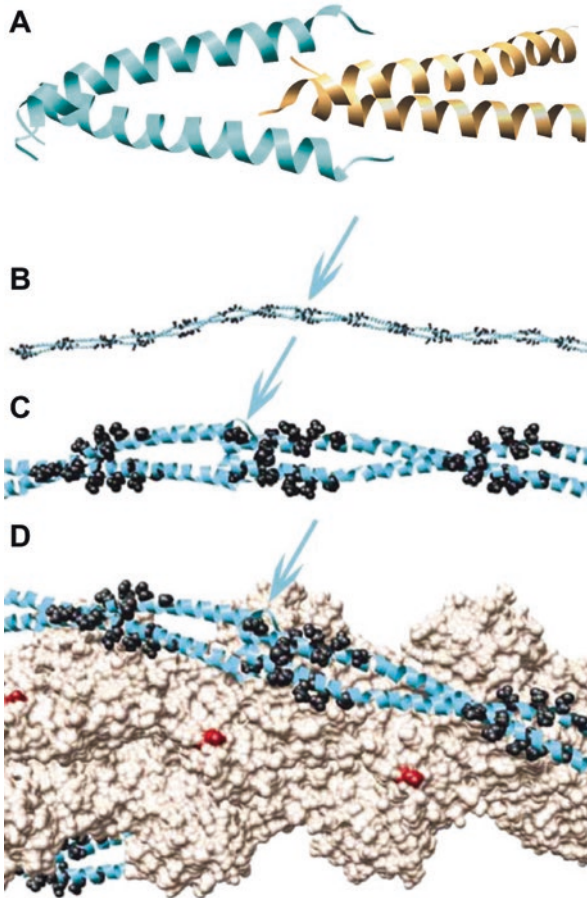
Fig. 9.3 Solution structure of the tropomyosin overlap complex and relationship of actin binding sites to the actin filament. (**A**) Ribbon diagram of the NMR structure of the striated muscle tropomyosin overlap complex (PDB ID 2G9J). (**B**) Two molecules of full-length tropomyosin molecules from the 7 Å structure (1C1G) with side-chains of the Phillips consensus actin binding sites shown in *black* (Phillips 1986). The *arrow* points to the site of the junction. (**C**) Detail of the junction between the two molecules in Fig. 9.3B shows three contiguous actin binding sites. (**D**) Model of the joined tropomyosins on a model of the actin filament. Residue K238 of actin is shown in *red* as a position marker. (**C**) and (**D**) show the 90° relationship of the two coiled coils in the overlap complex result in the actin binding sites on either side of the junction having a similar relationship to actin (Figure modified from Greenfield et al. (2006))

Å in some cases, which is the length of one or more actin subunits in the filament, and therefore may be significant for regulation (Brown 2013). Hodges and his colleagues have suggested that specific regions control long-range stability of tropomyosin without being responsible for folding *per se* (Kirwan and Hodges 2014).

Actin binding and regulatory functions of tropomyosin depend on cooperativity and synergy. Intramolecular synergy is recognized to be widespread as we appreci-

ate that coiled coils are more than dimerization motifs. For example, in dynein a 10 nm coiled coil (the "stalk") links the microtubule binding domain with the ATPase domain and enables two-way communication between them. Models have been put forward, but the details of the mechanism remain unresolved (Carter and Vale 2010; Nishikawa et al. 2015). Mechanisms of transmission of structural information in coiled coils are likely to be relevant to a variety of cytoskeletal functions, as well as communication in trans-membrane helices.

9.3 How Does Tropomyosin Bind to Actin? A Design for Weak and Dynamic Binding to the Actin Filament

The actin filament is the universal target of tropomyosin. The association of tropomyosin with actin was first suggested based on co-extraction from muscle (Corsi and Perry 1958), and later shown to be specific (Laki et al. 1962). Hanson and Lowy proposed a model of the actin filament based on negatively-stained EM images of actin, and suggested that there are two strands of tropomyosin that run in the two grooves of the actin helix (Hanson and Lowy 1963). This proposal was supported by the first three-dimensional reconstructions of electron micrographs of thin filaments (Moore et al. 1970). A series of thin filament structures followed, most recently with near-atomic resolution of actin using cryo-electron microscopy in which actin can be modeled using crystal structures of monomeric actin (Behrmann et al. 2012; Galkin et al. 2015; Sousa et al. 2013; von der Ecken et al. 2015).

Tropomyosin binds end-to-end to form cables along both sides of the helical filament such that the N terminus is oriented towards the pointed, slow-growing end of the filament (towards the M-line in striated muscle sarcomeres). The position tropomyosin occupies on the actin filament depends on the regulatory state (Fig. 9.4A). In the context of the steric blocking model (see Chap. 11; Paul et al. (2017)), with troponin in the absence of Ca^{2+}, tropomyosin occupies the myosin binding site on actin at the periphery of the helix, thus inhibiting myosin from binding to actin. With troponin in the presence of Ca^{2+}, and when the regulated actin filament is fully activated by myosin with troponin-Ca^{2+}, tropomyosin progressively shifts (slides or rolls) towards the center of the actin helix to free the myosin binding sites (Lehman et al. 2013). In the absence of troponin and myosin, the position of tropomyosin is variable, located toward the periphery of the helix, and can depend on the tropomyosin isoform (Lehman et al. 2000; Sousa et al. 2013; von der Ecken et al. 2015). The recent cryo-EM structures with near-atomic resolution reveal the average azimuthal position of tropomyosin on the actin filament, a significant advance. The helical averaging used in the helical reconstructions, and the lower resolution of tropomyosin than actin (tropomyosin, 6.5 Å; actin, 3.7 Å in (von der Ecken et al. 2015)) imposes limitations on the structural information about the relationship of tropomyosin to actin. Although the coiled coil helices are well resolved, the axial position of tropomyosin on the filament (where the ends are), the rotational position
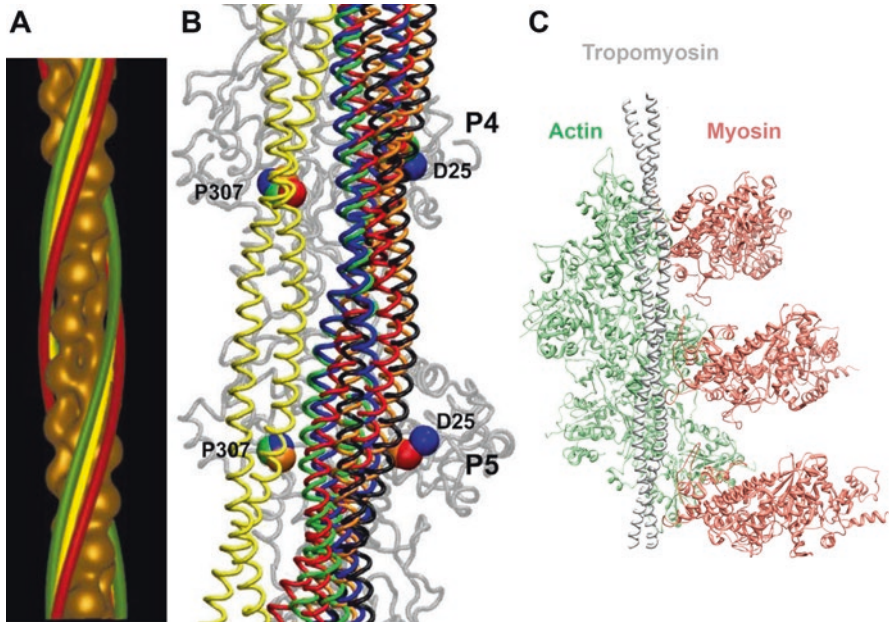
Fig. 9.4 Models of the thin filament showing the different regulatory states. (**A**) The three positions of tropomyosin on actin. *Red*, blocked; *yellow*, closed; *green*, open; (Modified with permission from Craig and Lehman (2001)). (**B**) Average structures of periods 4 and 5 of tropomyosin bound to actin in comparison with tropomyosin in the closed and open positions. The average structures of tropomyosin from four aMD simulations are colored *blue*, *red*, *green* and *orange*. The tropomyosin model at the closed position is colored *black*. The tropomyosin model in the open position is colored *yellow* (Modified from Zheng et al. (2013)). (**C**) A model of actin-tropomyosin-myosin S1 showing the proximity of tropomyosin to myosin (PDB ID 4A7F) (Behrmann et al. 2012)

(which surface faces the actin filament), variations in the coiled coil structure (discussed in Sect. 9.2), and small variations in tropomyosin-actin along the length of tropomyosin are lost in averaged structures. MD simulations coupled with optimization of critical electrostatic interactions have proved to be powerful tools for development of plausible atomic models of the actin-tropomyosin filament (Behrmann et al. 2012; Brown et al. 2005; Fischer et al. 2016; Li et al. 2011; Orzechowski et al. 2014; Sousa et al. 2013; von der Ecken et al. 2015; Zheng et al. 2013).

To understand how tropomyosin binds actin requires a different way of thinking than for high affinity complexes between globular proteins. Tropomyosin (center of the coiled coil) is 38–40 Å distant from the central axis of the actin filament in the most recent structure (von der Ecken et al. 2015). This is considered to be too far for extensive formation of salt bridges or strong hydrophobic interactions, though molecular dynamics simulations have identified locations where ion pairs can form between tropomyosin and actin, and where residues are close enough to actin to form Van der Waals contacts. The predominant weak, long-distance interactions

with the actin filament enable tropomyosin to shift between regulatory states on the surface of the actin filament with a low energy cost. Phillips described it as a "restless cable" on the filament (Phillips et al. 1986). A molecular dynamics simulation starting with the EM-based actin-tropomyosin model available at the time (Li et al. 2011) showed a considerable extent of fluctuation of tropomyosin on the surface of the actin filament, with individual periods fluctuating variably between positions (Fig. 9.4B) (Zheng et al. 2013). Tropomyosin is, in essence, a continuous semi-flexible cable that floats on the surface of the actin filament enabling a rapid, cooperative response to signals, such as Ca^{2+} binding to troponin in striated muscle, and myosin binding to the actin filament. Consistent with this picture is the general failure of tropomyosin to be chemically crosslinked to actin. So, if we can't "see" actin binding sites, how do we view them?

9.3.1 What Constitutes an Actin Binding Site on Tropomyosin?

Tropomyosin binds weakly to actin, with a ~mM K_d (Wegner 1980). The apparent dissociation constant in the μM or sub-μM range (depending on the isoform and conditions) depends on cooperativity enabled by the end-to-end association of tropomyosin on actin (Wegner 1980). Binding mutually stabilizes both tropomyosin and the actin filament (Ishii and Lehrer 1985; Levitsky et al. 2000; Singh and Hitchcock-DeGregori 2009; Wegner 1982; Weigt et al. 1991). A combination of computational, experimental, and structural approaches has resulted in a working model for actin binding sites on tropomyosin that has been incorporated into current models.

Striated muscle tropomyosin spans the length of seven actins in the filament. In a sequence analysis that takes into account the coiled coil structure, Phillips noted a sevenfold pattern of charged and hydrophobic surface residues that he proposed were quasi-equivalent actin binding sites (Fig. 9.5A) (Phillips 1986). Recombinant DNA methodology opened the way to test the model, and other structural hypotheses. In summary:

1. The periodic repeats do function as quasi-equivalent actin binding sites (Singh and Hitchcock-DeGregori 2007).
2. Certain of the seven quasi-equivalent repeats are more critical for actin binding and regulatory function than others (Hitchcock-DeGregori and An 1996; Hitchcock-DeGregori et al. 2001; Kawai et al. 2009; Landis et al. 1997).
3. A model-independent evolutionary approach focused on the role of conserved surface residues in the seven tropomyosin periods revealed a pattern of acidic and basic residues in the first halves of the seven repeats, that contribute to actin affinity; some periods are more important for actin binding than others (Fig. 9.5B) (Barua et al. 2011, 2013).
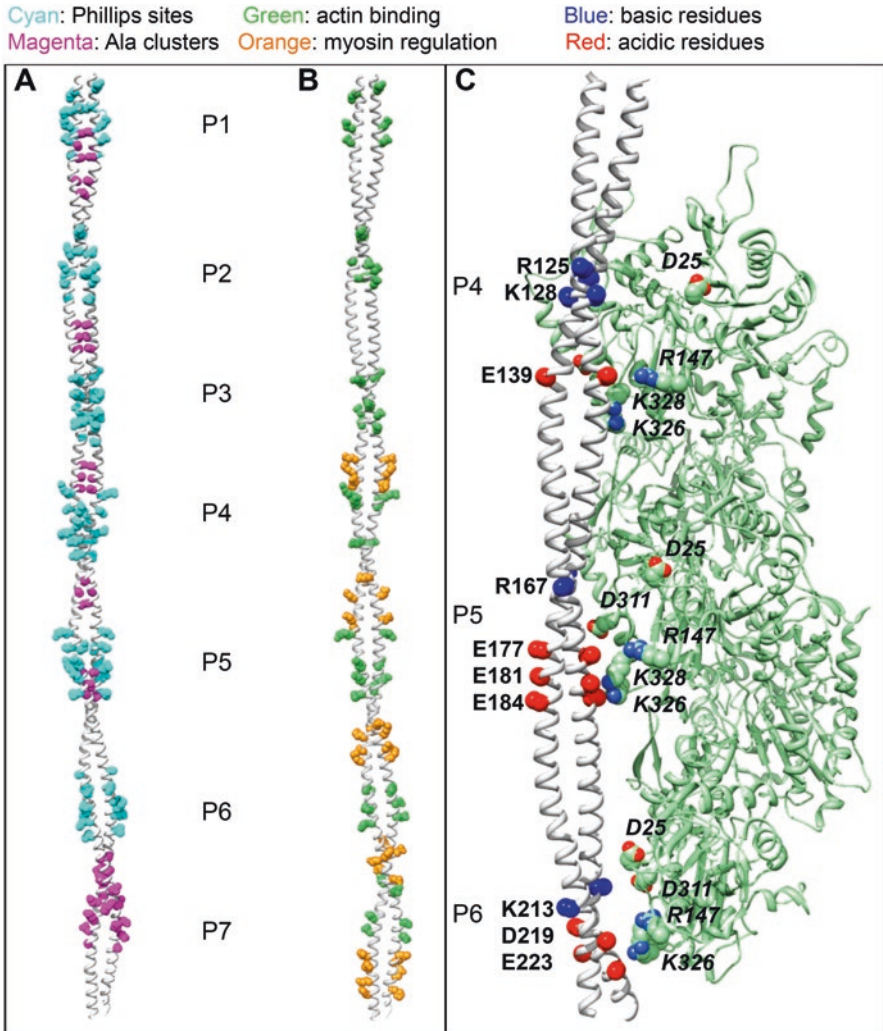
Fig. 9.5 Structure of tropomyosin showing periodic repeats, free and in relation to actin. The N terminus is at the top of the figure. (**A**) The 7 Å structure (PDB ID 1C1G) showing the Phillips consensus sites in P1-P6 in cyan and seven residues in P7 displayed in magenta (Phillips 1986) and destabilizing interface AlaSer clusters in magenta. (**B**) The evolutionarily-conserved actin binding sites on tropomyosin are colored *green*; evolutionarily-conserved residues important for regulating myosin are in *orange* (Barua et al. 2012, 2013). (**C**) A model of periods 4 and 5 of tropomyosin on actin illustrates the proximity of evolutionarily conserved actin binding site residues to a proposed binding site on actin. (PDB ID 3J8A) (von der Ecken et al. 2015)

4. The Ala clusters that destabilize the coiled coil are most critical when localized within or proximal to an actin binding site suggesting a possible role for intrinsically disordered regions in binding actin (Graceffa and Lehrer 1980; Ishii et al. 1992; Ishii and Lehrer 1985; Singh and Hitchcock-DeGregori 2006, 2009).
5. Models of tropomyosin on actin primarily based on optimization of critical electrostatic interactions (Barua et al. 2013; Brown et al. 2005) and more recent molecular dynamics simulations (Li et al. 2011) identified a putative tropomyosin binding site on subdomains 1 and 3 of actin (Fig. 9.5C). Mutation of actin residues in the proposed site resulted in weakened actin affinity, lending support to the model (Barua et al. 2013).
6. A consequence of the 90° rotation of the ends in the overlap complex is that the orientation of equivalent residues on tropomyosin, such as those involved in binding actin, have equivalent positions in relation to actin in successive tropomyosin molecules (Fig. 9.3B-D) (Greenfield et al. 2006), supported by a molecular dynamics model (Orzechowski et al. 2014). A further consideration is that some tropomyosins exist as heterodimers of different polypeptide chains, providing an additional source of thin filament diversity.

In our view it is misleading to think of tropomyosin as having a discrete structure or location on the filament. The interactions, primarily electrostatic, are simply too weak. The bends, staggers, holes, and a flexible overlap complex afforded by diversions from canonical coiled coil structure give tropomyosin a degree of directional flexibility needed to bind to the helical actin filament and change position on the actin helix in response to binding of myosin and other actin filament regulatory proteins. The staggers and other perturbations of the coiled coil may carry structural information long range in the molecule, important for cooperative regulation (Brown 2013). Overall the results in the literature support Brown's view that tropomyosin is semi-flexible, that it is not pre-formed to bind to the actin helix as the "Gestalt" model proposes (Brown 2010; Holmes and Lehman 2008). To move forward we need to know the structure and dynamics of tropomyosin at atomic resolution in solution, free and bound to the actin filament, work that should soon be feasible using solid state NMR. In the meantime, advances in molecular dynamics simulations offer a powerful way to understand the structure and dynamics of full-length tropomyosins free, and in macromolecular assemblies, and the possible consequences of disease-causing mutations.

9.3.2   How Does Tropomyosin Assemble on the Actin Filament?

Tropomyosin binding to actin is rapid and is thought to initiate at isolated sites that then extend in a contiguous manner (Tobacman 2008; Weigt et al. 1991). Is the binding initiated at the pointed end of the filament or at a random site(s) along the filament (Hsiao et al. 2015; Schmidt et al. 2015)? The cooperative signal may be

transmitted through the tropomyosin via the overlap complexes between the ends (Geeves and Lehrer 2002; Lehrer and Geeves 1998) or through or across the actin filament (Butters et al. 1993; Moraczewska and Hitchcock-DeGregori 2000). These questions have not been definitively answered; answering them is central to understanding the dynamics of actin filaments in cells.

## 9.4	Regulation of the Actin-Myosin Interaction

So far, we have considered how tropomyosin binds actin, and the cooperativity of binding. What does tropomyosin do to the actin filament? What function does it serve? One is the cooperative regulation of myosin interaction with actin. Muscle contraction and numerous intracellular motile processes depend on the binding of myosin to actin and on actin activation of ATP hydrolysis by myosin. Actin accelerates product release (the rate limiting step) following ATP hydrolysis, and this step is in turn regulated by tropomyosin.

It is easy to understand the significance of cooperative regulation when considering muscle contraction. When a skeletal muscle fiber is activated, the Ca^{2+} released from the sarcoplasmic reticulum binds to troponin and cooperatively switches the actin filament from the relaxed to the active state in a matter of milliseconds. Both Ca^{2+} binding to troponin and myosin binding to actin are required for full activation (Gordon et al. 2000; Moore et al. 2016). As put forth in the steric blocking model (see Chap. 11; Paul et al. (2017)), in the absence of myosin, tropomyosin binds to actin at or near the myosin binding site, and myosin binding to actin shifts the tropomyosin azimuthally to the "open" site with full activation of the filament (Fig. 9.4).

The story about the role of tropomyosin in regulation begins with two landmark papers. Weber and her colleagues showed that at low MgATP concentration (the substrate of myosin), myosin heads activate a regulated actin filament (actin-tropomyosin-troponin) and that tropomyosin alone enables a cooperative response (Bremel and Weber 1972). In an independent study, Eaton showed that tropomyosin increases the affinity of myosin for actin and, conversely, that myosin increases the affinity of tropomyosin for actin (Eaton 1976). Lehrer and colleagues then showed that myosin binding to actin-tropomyosin could cooperatively switch the actin filament from an inactive to a fully active state with substoichiometric occupancy by myosin (Ishii and Lehrer 1985; Lehrer and Ishii 1988). A recent single molecule study suggests that local activation of a regulated actin filament requires at least two myosin heads, but the mechanism and size of the activated region are unresolved (Desai et al. 2015). Until recently, tropomyosin has been considered a passive player in this ballet. Might tropomyosin *directly* affect myosin?

Tropomyosin can inhibit or activate myosin function, depending on the isoforms of both proteins, with vertebrate as well as yeast proteins (Table 9.2) (Barua and Hitchcock-DeGregori 2015; Barua et al. 2012, 2014; Chacko 1981; Clayton et al. 2010, 2013; Fanning et al. 1994; Hodges et al. 2012; Hundt et al. 2016; Lehrer and Morris 1982, 1984; McIntosh et al. 2015; Sckolnick et al. 2016; Tang and Ostap 2001;

Table 9.2 Effect of tropomyosin on actomyosin function

Myosin	Tropomyosin	Effect on myosin	Assay[1, 2, 3, 4]	References
Vertebrate muscle				
Skeletal	Tpm1.1/2.2 (striated)	Inhibition	1	Eaton et al. (1975); Lehrer and Morris (1982)
	Tpm1.1	Inhibition	2, 3	Barua et al. (2012)
		Inhibition	1	Fanning et al. (1994)
		No effect	2	Fanning et al. (1994)
	Tpm1.3/2.1 (smooth)	Activation	1	Lehrer and Morris (1984)
	Tpm1.6	Inhibition	1	Fanning et al. (1994)
		No effect	2	Fanning et al. (1994); Barua et al. (2014)
	Tpm1.8	Activation	2	Barua et al. (2014)
	Tpm1.7	Inhibition	1	Fanning et al. (1994)
		No effect	2	Fanning et al. (1994)
	Tpm1.9	Activation	1, 2	Fanning et al. (1994)
	Tpm3.1	Activation	2	Barua et al. (2014)
		No effect	3	Barua et al., unpublished[1]
	Tpm4.2	No effect	2	Barua et al. (2014)
Smooth	Tpm1.3/2.1	Activation	1, 2	Chacko (1981);
	Tpm1.1/2.2	Activation	1	Umemoto et al. (1989); Yamaguchi et al. (1984)
Vertebrate non-muscle				
Myosin I	Tpm1.9	Inhibition	1, 2	Fanning et al. (1994)
	Tpm1.7	Inhibition	1, 2	Fanning et al. (1994)
	Tpm1.1	Inhibition	1, 2	Fanning et al. (1994)
	Tpm1.6	Inhibition	1, 2	Fanning et al. (1994), Kee et al. (2015) and Tang and Ostap (2001)
	Tpm3.1	Inhibition	2	Kee et al. (2015)
Myosin IIA	Tpm1.1	No effect	2	Barua et al. (2014)
	Tpm1.6	No effect	2	Barua et al. (2014)
	Tpm1.8	Activation	2	Barua et al. (2014)
	Tpm3.1	Activation	1, 2	Barua et al. (2014)
	Tpm4.2	Activation	1, 2, 3	Barua et al. (2014); Hundt et al. (2016)
Myosin Va	Tpm1.8	Activation	4	Sckolnick et al. (2016)
	Tpm4.2	Inhibition	4	Sckolnick et al. (2016)
	Tpm3.1	Activation	4	Sckolnick et al. (2016)

(continued)

Table 9.2 (continued)

Myosin	Tropomyosin	Effect on myosin	Assay[1,2,3,4]	References
Saccharomyces cerevisiae				
Myo2p	Tpm1p	Activation	4	Hodges et al. (2012)
(Myosin V)	Tpm2p	Activation	4	Hodges et al. (2012)
	Tpm1.3/2.1	Activation	4	Hodges et al. (2012)
Schizosaccharomyces pombe				
Myo1p	Cdc8p	Inhibition	1, 2	Clayton et al. (2010)
Myo51p	Cdc8p	Activation	1	Clayton et al. (2010)
		Inhibition	2	Clayton et al. (2010)
Myo52p	Cdc8p	Activation	1, 4	Clayton et al. (2013);
	Cdc8p	Inhibition	2	Clayton et al. (2014); Clayton et al. (2010)
Myo2p	Cdc8p	Activation	1	Stark et al. (2010)
		Inhibition	2	Stark et al. (2010)

[1]Steady-state ATPase

[2]*In vitro* motility assay

[3]Rate of Pi release (stopped-flow)

[4]Single molecule TIRF microscopy

Umemoto et al. 1989; Williams et al. 1984). Tropomyosin can enable myosin to move faster, or slower on actin, activating or inhibiting ATP hydrolysis, or allow a processive myosin to stay on the actin filament for a shorter or longer time. In most cases, the results are interpreted in terms of the position of tropomyosin on actin and the ease with which myosin, by binding actin, can switch the position of tropomyosin to "open" on the actin filament, thereby vacating the actin sites for strong binding of myosin and force development (Fig. 9.4). This view has persisted, despite evidence that tropomyosin can directly interact with myosin and be crosslinked to myosin on actin (Golitsina and Lehrer 1999; Merkel et al. 1989; Tao and Lamkin 1984). Interest in a possible direct interaction between tropomyosin and myosin has been recently rekindled with the models of a near-atomic resolution structure of actin-tropomyosin-myosin S1, that illustrate the proximity of tropomyosin to myosin (Fig. 9.4C) (Barua et al. 2012; Behrmann et al. 2012).

Here we put forward the hypothesis that *direct* myosin regulation is a core, conserved function of tropomyosin, along with actin binding. We further postulate that the diversity of tropomyosins (more than 40 isoforms in vertebrates) at least in part relates to regulation of specific myosins, a primary physiological role. We discuss elements in tropomyosin that have universal and specific roles for regulating the function and possibly the affinity for and cellular localization of myosins to specialized actin tracks within cells. As a multiplicity of myosin isoforms becomes available for cellular and biophysical studies, we anticipate attention will turn to deciphering mechanisms of regulation at the level of the actin filament, the cellular

machinery at the terminus of cell signaling cascades. Tropomyosin may regulate myosin at two levels:

1. By directly or indirectly regulating actin activation of ATP hydrolysis by myosin (by accelerating product release or another kinetic step) and
2. By physically acting as a tether to increase the affinity of myosin for actin or influencing force transduction in an isoform-specific way.

9.4.1 Common Features of Regulation

How is tropomyosin designed to regulate myosin? Is there a conserved "code" as there is for actin binding (Fig. 9.5) (Barua et al. 2011, 2013)? A mutagenesis approach allowed identification of conserved surface residues in the second halves of the periodic repeats, especially periods 3 and 6, which regulate the activation of skeletal muscle myosin by striated muscle actin-tropomyosin (Barua et al. 2012). The residues are C-terminal to the actin binding sites, illustrated in Fig. 9.5B. A similar pattern persists in a non-muscle tropomyosin, which is encoded by a different gene for the regulation of skeletal muscle myosin (Barua and Hitchcock-DeGregori 2015) and vertebrate myosin Va[1]. We propose that these conserved sites may regulate the actin-myosin interaction via a common mechanism that may include direct interaction between myosin and tropomyosin (Fig. 9.4C).

9.4.2 Specific Regulation

The isoform specificity of myosin regulation indicates that an actin-tropomyosin filament can be attractive to one myosin and repulsive to another. In general terms, the effects can be considered in relation to the affinity of myosin for actin-tropomyosin (Table 9.2). While differences between striated and smooth muscle tropomyosins and myosins have long been known, recent work indicates that a finely tuned specificity underlies both muscle and non-muscle myosin regulation. For example, a non-muscle isoform, Tpm4.2 (encoded by the TPM4 gene, Fig. 9.6) activates a non-muscle filamentous myosin, NMIIA, has no effect on skeletal myosin II, but inhibits a processive, cargo-carrying myosin Va (Barua et al. 2014; Hundt et al. 2016; Sckolnick et al. 2016). Tpm3.1, another non-muscle tropomyosin encoded by the TPM3 gene, activates all three myosin isoforms (Barua et al. 2014; Sckolnick et al. 2016). Some of the effects involve common mechanisms (see Sect. 9.4.1), but in other aspects tropomyosin is designed to regulate one myosin, but not another. We will discuss this in Sect. 9.5 on tropomyosin isoform specificity.

---

[1] Barua, B., Sckolnick, M., White, H.D., Trybus, K.M., Hitchcock-DeGregori, S.E., unpublished results
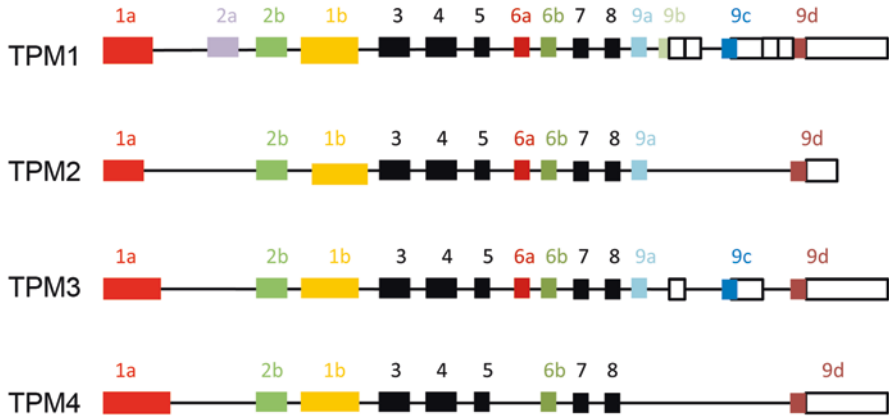
Fig. 9.6 Gene structure of the human tropomyosin genes. Most vertebrates have four tropomyosin genes, each encoding multiple isoforms due to alternate exon expression. Alternate promoters select between exon 1a and 2 to encode ~284 residue tropomyosins (HMW), or exon 1b that encode ~247 residue tropomyosins (LMW) that lack an exon 2. Three exons are alternatively spliced (2a/2b, 6a/6b, 9a/9b/9c/9d). The *black* bars are constitutively expressed (Figure provided by M. Geeves. Geeves et al. (2015))

9.5 Four Tropomyosin Genes, Many Isoforms and Diverse Binding Partners

Most of our discussion has been of striated muscle tropomyosin, the form that has been most extensively studied and for which we have structural information. This belies the complexity and diversity of tropomyosin functions, as we consider that mammalian cells have the potential of expressing ~40 distinct isoforms and even more with the possibility for heterodimer formation. Cells typically express a subset of potential isoforms (Gunning et al. 2008; Schevzov et al. 2011). We limit our discussion here to mammalian tropomyosins. For yeast tropomyosins and their functions, we refer the reader to (Kovar et al. 2011; Pruyne 2008). An additional "silent" source of diversity, that remains to be explored, is the relationship to different actin isoforms.

In mammals, four genes encode tropomyosin: TPM1, TPM2, TPM3, and TPM4 (Geeves et al. 2015). In Fig. 9.6, the colors indicate exons that are alternate (a-d) in one or more genes; exons in black are constitutive, expressed in all isoforms. Alternative promoters select between expression of exon 1a + 2a/2b, or exon 1b. This results in high molecular weight forms (HMW, ~284 amino acids per chain that span seven actin subunits in the filament) with exons 1a + 2, and low molecular weight forms (LMW, ~247 amino acids per chain that span six actin subunits) that begin with exon 1b and lack an exon 2 that corresponds to one periodic repeat. Additional diversity results from alternative splicing of one or more exons, depending on the gene: 2a/2b, 6a/6b, 9a/9b/9c/9d. The most variable regions of tropomyosin are the ends, encoded by alternate forms of the 1st and 9th exons that form the

intermolecular "overlap" junction (Fig. 9.3A). The orientation of tropomyosin on the actin is such that the C terminus (encoded by the ninth exon) is oriented towards the barbed, fast-growing end of the filament, located in Z-lines in striated muscle sarcomeres and the cell membrane in filopodia and focal adhesions. The N terminus (encoded by the first exon) is oriented towards the pointed, slow-growing end of the actin filament (towards the M-line in striated muscle sarcomeres) where it can be capped by tropomodulin (Colpan et al. 2013). Tropomyosin alone can affect the dynamics at the filament ends in an isoform-specific way (Khaitlina 2015).

Tropomyosin isoforms differ in actin affinity and in cellular functions (Gunning et al. 2015; Khaitlina 2015; Wang and Coluccio 2010). They have preferential localizations in regions of the stress fibers, lamellipodia, and filopodia, depending on the cell type (Schevzov et al. 2011), as well as different roles in actin assemblies (Tojkander et al. 2011). Specific tropomyosins are associated with stages of embryonic development and morphogenesis, cellular mechanics and migration, cytokinesis, intracellular transport, metabolism and cancer, with much to be discovered and understood (reviewed in (Gunning et al. 2015; Vindin and Gunning 2013)). Even though tropomyosin isoforms differ in actin affinity, it seems likely that the specificity primarily relates to functions with other muscle and cytoskeletal proteins rather than to serve as mechanisms of redundancy.

The best-documented cellular example of an isoform-specific function is the change in tropomyosin expression that takes place upon oncogenic transformation (reviewed in Choi et al. (2012); Gross (2013); Helfman et al. (2008)). Down-regulation of HMW tropomyosins is commonly associated with the morphological changes in transformed cells and tumors (Hendricks and Weintraub 1981; Raval et al. 2003). Overexpression of a HMW tropomyosin restores the presence of stress fibers and anchorage-dependent cell growth in cells (Mahadev et al. 2002; Prasad et al. 1993, 1999). The phenomenon is widespread in different cell lines and tumors. Studies indicate that LMW forms, in particular isoforms encoded by the TPM3 gene, are critical for tumor survival, offering potential for the design of drugs that interfere with LMW tropomyosin function (Bonello et al. 2016; Stehn et al. 2013). More generally, these studies point towards isoform-specific regulation of actin filament function and dynamics within cells, and the regulation of the interaction of filaments with focal adhesions, where tropomyosin may influence cellular migration and and/or the rigidity of the extracellular matrix (Wolfenson et al. 2016).

The molecular basis of isoform specificity and function is a frontier of research that will lead towards understanding the breadth and subtlety of tropomyosin regulation of contraction, motility and cytoskeletal dynamics. *In vivo* and live cell studies have been a challenge because of the complex pattern of isoform expression. The development of specific antibodies has enabled characterization of the tissue-specificity and localization of many of these isoforms (Schevzov et al. 2011). The critical function of the ends of tropomyosin compromises and challenges the use of C-terminal or N-terminal fluorescent protein tags, such as GFP, or epitope tags. In the future, application of CRISPR/Cas9 methodology (Vandemoortele et al. 2016) to introduce carefully designed probes, coupled with super-resolution light microscopy may allow analysis of the localization and dynamics of tropomyosin isoforms in living cells. In the meantime, much has been learnt from *in vitro* studies.

In addition to binding actin filaments and regulating myosin as previously discussed, tropomyosin interacts directly or indirectly with proteins that regulate actin filament polymerization and assembly into higher order structures. In many cases the regulation is isoform specific. Interacting proteins that regulate muscle contraction include troponin (Gordon et al. 2000), caldesmon (Wang and Coluccio 2010), and possibly nebulin that runs the length of the actin filament (Castillo et al. 2009; Pappas et al. 2011). In a new twist, the thick filament protein, myosin-binding protein C displaces tropomyosin on the thin filament (Mun et al. 2014). As a universal regulator of the dynamics and assembly of the actin cytoskeleton, tropomyosin alone can inhibit polymerization and depolymerization at the filament ends as well as bind with isoform-specificity to proteins that regulate the dynamics of the filament ends (Carlier et al. 2015; Khaitlina 2015). These include formin, a protein that promotes polymerization at the barbed end of the actin filament (Carlier et al. 2015; Wawro et al. 2007). The other end, the pointed end, is often capped in muscle and non-muscle cells by tropomodulin, a family of proteins that bind to actin and tropomyosin, with isoform specificity (Colpan et al. 2013). Tropomyosin isoforms differ in their ability to protect the filament from severing and pointed end depolymerization (Khaitlina 2015).

By binding to proteins that regulate actin polymerization and depolymerization and by competing with actin severing and certain crosslinking proteins, tropomyosin has a major influence on the local organization and function of actin filaments. For example, actin filaments form a branched, dendritic network at the edge of many migrating cells that depends on nucleation of new filaments from the sides of actin filaments by the Arp2/3 complex. Together with severing and pointed-end depolymerization of actin filaments by ADF/cofilin a network forms at the periphery of the cell with actin filaments that grow at the barbed end to drive the cell forward (Blanchoin et al. 2014). Tropomyosin is an isoform-specific antagonist to Arp2/3 complex, cofilin, and the membrane-bound myosin class I proteins (Bravo-Cordero et al. 2013; Khaitlina 2015; McIntosh et al. 2015; Ostap 2008; Robaszkiewicz et al. 2015). A consequence is that filaments in dendritic actin networks are tropomyosin-poor (DesMarais et al. 2002).

How can we decipher isoform specificity in a continuous coiled coil in which domains cannot be visualized as independently folded regions of the protein? The highly conserved exon structure of the tropomyosin genes indicates that tropomyosin is a functionally modular protein. In instances where isoform specificity has been investigated, the alternatively expressed exon 1 and exon 9 that encode the N-terminal and C-terminal domains are key.

9.5.1   Exon 9 Encodes Isoform-Specific Function

Findings on exon 9 illustrate examples of isoform specificity. Exon 9 (Fig. 9.6) encodes the C-terminal 27 residues of tropomyosin, C-terminal to the last actin binding site (Fig. 9.5A, B). The last 11–14 residues complex with the N terminus

(Fig. 9.3A), but the first part of the exon encodes isoform-specific function. Troponin T, a component of troponin, the Ca^{2+}-regulatory protein in striated muscle, binds to exon 9a in striated muscle tropomyosin (a striated muscle-specific exon), but not to proteins containing exon 9d sequences (Hammell and Hitchcock-DeGregori 1996; Palm et al. 2003). Mutation of surface residues in this region of striated muscle tropomyosin or the widely expressed non-muscle tropomyosin, Tpm3.1, have no effect on muscle myosin function, inferring that skeletal muscle myosin has no interaction with exon 9-encoded sequences. In contrast, analogous exon 9 mutations in Tpm3.1 inhibit the processivity of the non-muscle cargo-carrying myosin Va, suggesting an isoform-specific role in defining the actin track for that cellular myosin[1]. These results support the view that a role of tropomyosin is to specify actin filaments in cells for their specific motile and structural functions, in some cases possibly by recruiting myosin to a class of actin filaments in the cytoskeleton.

9.5.2   Complexes Between Tropomyosin and Tropomodulin

At the pointed end of actin filaments, the N terminus of tropomyosin is free and not in complex with the C terminus as it is along the length of the actin filament. There the N-terminal region of the capping protein, tropomodulin binds the N termini of the tropomyosin molecules on the two sides of the actin filament, with an intervening actin binding site (Fig. 9.7A) (Kostyukova 2008b). The binding to tropomyosin, in a region encoded by exon 1, is isoform specific (Colpan et al. 2013; Yamashiro et al. 2012). What might such a complex look like? We show here conceptual models based on biophysical and structural studies of tropomyosin binding to tropomodulin at the end of the actin filament (Fig. 9.7B-D) (Kostyukova et al. 2005, 2006, 2007; Rao et al. 2014). Circular dichroism and NMR studies have shown that the tropomyosin-binding domains of tropomodulin become more helical upon binding tropomyosin, stabilizing both proteins (Kostyukova et al. 2005, 2007). The results are consistent with the formation of helix bundles, supported by modeling and mutagenesis studies (Fig. 9.7B, C). Crystal structures of the two actin binding sites of tropomodulin on actin have been determined, and are proposed to bind to sites in different actin protomers at the pointed end of the filament (Fig. 9.7D) (Rao et al. 2014).

The tropomyosin-tropomodulin models in Fig. 9.7 allow appreciation of the pleitropic nature of tropomyosin. The different tropomodulin binding sites on the tropomyosins on the two sides of the actin filament, together with the binding of tropomodulin to actin subunits in different protomers, make the pointed end of the actin filament asymmetric. The asymmetry and tropomodulin-tropomyosin complexes may have long-range effects and functional consequences. The isoform specificity of tropomyosin-tropomodulin interactions adds a mechanism to create actin filament diversity within cells (Colpan et al. 2013; Yamashiro et al. 2012).
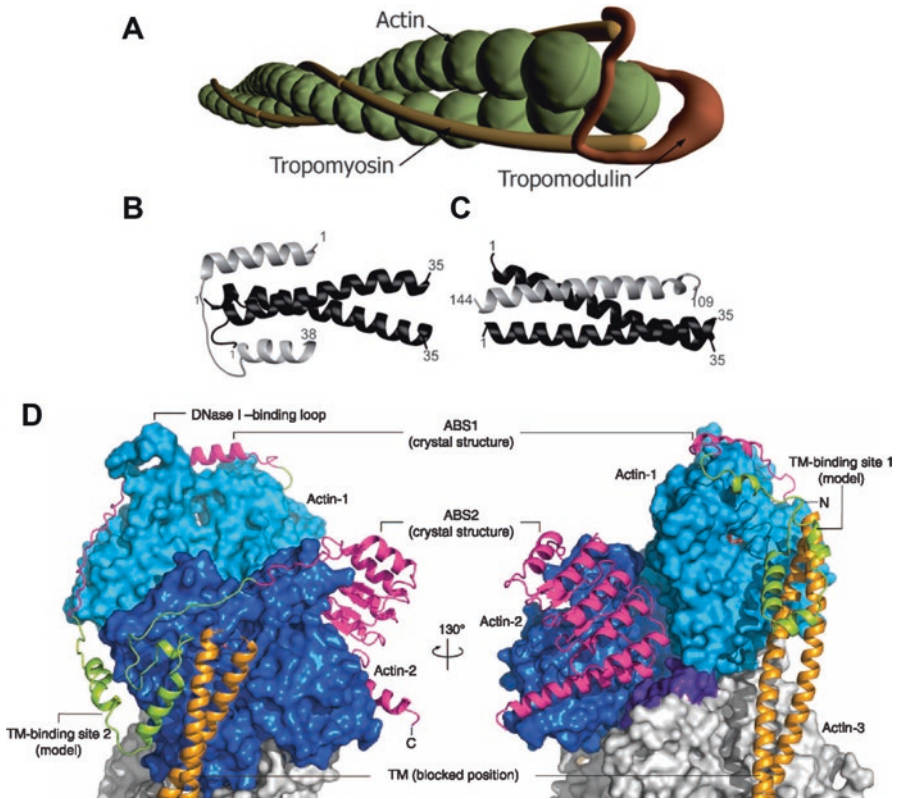
Fig. 9.7 Models of the pointed end of the actin filament with tropomyosin and the pointed-end capping protein, tropomodulin. (**A**) A cartoon representation of the actin-tropomyosin filament with a single tropomodulin binding the N termini of two tropomyosin molecules and both actin protomers in the filament. (**B**) and (**C**) Models of the two tropomyosin binding domains of Tmod1 binding to the N terminus of tropomyosin (LMW, encoding exon 1b form). (**A–C** modified with permission from Kostyukova (2008a) based on Kostyukova et al. (2007)). (**D**) Model of the pointed end based on x-ray structures, NMR structures, and peptide studies that have identified binding domains of tropomodulin. The model is based on the X-ray structures of Tmod1 peptides with the two actin binding sites (*magenta*) bound to the two actins in the protofilament (*blue*), incorporating models of Tmod1-muscle (HMW, encoding exon 1a) tropomyosin (Tmod1, *green*; tropomyosin, *gold*) into the filament (Modified with permission from Rao et al. (2014))

9.6   Conclusions

In this review we have given our perspective on the history of tropomyosin as it relates to its totally "modern" appearance with opportunities opening to understand how it regulates the actin filament in muscle and other cells. Much remains to be learned about the structure and dynamics of tropomyosin, its effect on the kinetics and mechanical properties of myosin, actin filaments, in single molecule studies,

within cells, and in interaction of cells with the extra cellular matrix. New and emerging methodologies will enable new investigators to unravel old mysteries and discover new ones.

References

Astbury WT, Reed R, Spark LC (1948) An X-ray and electron microscope study of tropomyosin. Biochem J 43:282–287

Bailey K (1946) Tropomyosin: a new asymmetric protein component of muscle. Nature 157:368

Bailey K (1948) Tropomyosin: a new asymmetric protein component of the muscle fibril. Biochem J 43:271–279

Bailey K (1951) End-group assay in some proteins of the keratin-myosin group. Biochem J 49:23–27

Bailey K, Gutfreund H, Ogston AG (1948) Molecular weight of tropomyosin from rabbit muscle. Biochem J 43:279–281

Barua B, Hitchcock-DeGregori SE (2015) Actomyosin regulation by conserved sites of Tm5NM1 (Tpm3.1), a nonmuscle tropomyosin. Biophys J 108(2):593a

Barua B, Fagnant PM, Winkelmann DA, Trybus KM, Hitchcock-DeGregori SE (2013) A periodic pattern of evolutionarily conserved basic and acidic residues constitutes the binding interface of actin-tropomyosin. J Biol Chem 288:9602–9609

Barua B, Nagy A, Sellers JR, Hitchcock-DeGregori SE (2014) Regulation of nonmuscle myosin II by tropomyosin. Biochemistry 53:4015–4024

Barua B, Pamula MC, Hitchcock-DeGregori SE (2011) Evolutionarily conserved surface residues constitute actin binding sites of tropomyosin. Proc Natl Acad Sci U S A 108:10150–10155

Barua B, Winkelmann DA, White HD, Hitchcock-DeGregori SE (2012) Regulation of actin-myosin interaction by conserved periodic sites of tropomyosin. Proc Natl Acad Sci U S A 109:18425–18430

Behrmann E, Muller M, Penczek PA, Mannherz HG, Manstein DJ, Raunser S (2012) Structure of the rigor actin-tropomyosin-Myosin complex. Cell 150:327–338

Blanchoin L, Boujemaa-Paterski R, Sykes C, Plastino J (2014) Actin dynamics, architecture, and mechanics in cell motility. Physiol Rev 94:235–263

Bonello TT, Janco M, Hook J, Byun A, Appaduray M, Dedova I, Hitchcock-DeGregori S, Hardeman EC, Stehn JR, Bocking T, Gunning PW (2016) A small molecule inhibitor of tropomyosin dissociates actin binding from tropomyosin-directed regulation of actin dynamics. Sci Rep 6:19816

Bravo-Cordero JJ, Magalhaes MA, Eddy RJ, Hodgson L, Condeelis J (2013) Functions of cofilin in cell locomotion and invasion. Nat Rev Mol Cell Biol 14:405–415

Bremel RD, Weber A (1972) Cooperation within actin filament in vertebrate skeletal muscle. Nat New Biol 238:97–101

Brown JH (2010) How sequence directs bending in tropomyosin and other two-stranded alpha-helical coiled coils. Protein sci 19:1366–1375

Brown JH (2013) Deriving how far structural information is transmitted through parallel homodimeric coiled-coils: a correlation analysis of helical staggers. Proteins 81:635–643

Brown JH, Cohen C (2005) Regulation of muscle contraction by tropomyosin and troponin: how structure illuminates function. Adv Protein Chem 71:121–159

Brown JH, Kim KH, Jun G, Greenfield NJ, Dominguez R, Volkmann N, Hitchcock-DeGregori SE, Cohen C (2001) Deciphering the design of the tropomyosin molecule. Proc Natl Acad Sci U S A 98:8496–8501

Brown JH, Zhou Z, Reshetnikova L, Robinson H, Yammani RD, Tobacman LS, Cohen C (2005) Structure of the mid-region of tropomyosin: bending and binding sites for actin. Proc Natl Acad Sci U S A 102:18878–18883

Butters CA, Willadsen KA, Tobacman LS (1993) Cooperative interactions between adjacent troponin-tropomyosin complexes may be transmitted through the actin filament. J Biol Chem 268:15565–15570

Carlier MF, Pernier J, Montaville P, Shekhar S, Kuhn S, Cytoskeleton D, Motility G (2015) Control of polarized assembly of actin filaments in cell motility. Cell Mol Life Sci 72:3051–3067

Carter AP, Vale RD (2010) Communication between the AAA+ ring and microtubule-binding domain of dynein. Biochem Cell Biol 88:15–21

Caspar DL, Cohen C, Longley W (1969) Tropomyosin: crystal structure, polymorphism and molecular interactions. J Mol Biol 41:87–107

Castillo A, Nowak R, Littlefield KP, Fowler VM, Littlefield RS (2009) A nebulin ruler does not dictate thin filament lengths. Biophys J 96:1856–1865

Chacko S (1981) Effects of phosphorylation, calcium ion, and tropomyosin on actin-activated adenosine 5′-triphosphatase activity of mammalian smooth muscle myosin. Biochemistry 20:702–707

Choi C, Kim D, Kim S, Jeong S, Song E, Helfman DM (2012) From skeletal muscle to cancer: insights learned elucidating the function of tropomyosin. J Struct Biol 177:63–69

Clayton JE, Sammons MR, Stark BC, Hodges AR, Lord M (2010) Differential regulation of unconventional fission yeast myosins via the actin track. Curr Biol 20:1423–1431

Clayton JE, Pollard LW, Sckolnick M, Bookwalter CS, Hodges AR, Trybus KM, Lord M (2013) Fission yeast tropomyosin specifies directed transport of myosin-V along actin cables. Mol Biol Cell 25:66–75

Clayton JE, Pollard LW, Sckolnick M, Bookwalter CS, Hodges AR, Trybus KM, Lord M (2014) Fission yeast tropomyosin specifies directed transport of myosin-V along actin cables. Mol Biol Cell 25:66–75

Cohen C, Caspar DL, Parry DAD, Lucas RM (1972) Tropomyosin crystal dynamics. Cold Spring Harb Symp Quant Biol 36:205–216

Colpan M, Moroz NA, Kostyukova AS (2013) Tropomodulins and tropomyosins: working as a team. J Muscle Res Cell Motil 34:247–260

Corsi A, Perry SV (1958) Some observations on the localization of myosin, actin and tropomyosin in the rabbit myofibril. Biochem J 68:12–17

Craig R, Lehman W (2001) Crossbridge and tropomyosin positions observed in native, interacting thick and thin filaments. J Mol Biol 311:1027–1036

Crick FHC (1953) The packing of alpha-helices: simple coiled-coils. Acta Cryst 6:9

Desai R, Geeves MA, Kad NM (2015) Using fluorescent myosin to directly visualize cooperative activation of thin filaments. J Biol Chem 290:1915–1925

DesMarais V, Ichetovkin I, Condeelis J, Hitchcock-DeGregori SE (2002) Spatial regulation of actin dynamics: a tropomyosin-free, actin-rich compartment at the leading edge. J Cell Sci 115:4649–4660

Eaton BL (1976) Tropomyosin binding to F-actin induced by myosin heads. Science 192:1337–1339

Eaton BL, Kominz DR, Eisenberg E (1975) Correlation between the inhibition of the acto-heavy meromyosin ATPase and the binding of tropomyosin to F-actin: effects of Mg2+, KCl, troponin I, and troponin C. Biochemistry 14:2718–2725

El-Mezgueldi M (2014) Tropomyosin dynamics. J Muscle Res Cell Motil 35:203–210

Fanning AS, Wolenski JS, Mooseker MS, Izant JG (1994) Differential regulation of skeletal muscle myosin-II and brush border myosin-I enzymology and mechanochemistry by bacterially produced tropomyosin isoforms. Cell Motil Cytoskeleton 29:29–45

Fischer S, Rynkiewicz MJ, Moore JR, Lehman W (2016) Tropomyosin diffusion over actin subunits facilitates thin filament assembly. Struct Dyn 3:012002

Frye J, Klenchin VA, Rayment I (2010) Structure of the tropomyosin overlap complex from chicken smooth muscle: insight into the diversity of N-terminal recognition. Biochemistry 49:4908–4920

Galkin VE, Orlova A, Vos MR, Schroder GF, Egelman EH (2015) Near-atomic resolution for one state of F-actin. Structure 23:173–182

Geeves MA, Lehrer SS (2002) Cooperativity in the Ca2+ regulation of muscle contraction. Results Probl Cell Differ 36:111–132

Geeves MA, Hitchcock-DeGregori SE, Gunning PW (2015) A systematic nomenclature for mammalian tropomyosin isoforms. J Muscle Res Cell Motil 36:147–153

Golitsina NL, Lehrer SS (1999) Smooth muscle alpha-tropomyosin crosslinks to caldesmon, to actin and to myosin subfragment 1 on the muscle thin filament. FEBS Lett 463:146–150

Gordon AM, Homsher E, Regnier M (2000) Regulation of contraction in striated muscle. Physiol Rev 80:853–924

Graceffa P, Lehrer SS (1980) The excimer fluorescence of pyrene-labeled tropomyosin. A probe of conformational dynamics. J Biol Chem 255:11296–11300

Greenfield NJ, Montelione GT, Farid RS, Hitchcock-DeGregori SE (1998) The structure of the N-terminus of striated muscle alpha-tropomyosin in a chimeric peptide: nuclear magnetic resonance structure and circular dichroism studies. Biochemistry 37:7834–7843

Greenfield NJ, Huang YJ, Palm T, Swapna GV, Monleon D, Montelione GT, Hitchcock-DeGregori SE (2001) Solution NMR structure and folding dynamics of the N terminus of a rat non-muscle alpha-tropomyosin in an engineered chimeric protein. J Mol Biol 312:833–847

Greenfield NJ, Swapna GV, Huang Y, Palm T, Graboski S, Montelione GT, Hitchcock-DeGregori SE (2003) The structure of the carboxyl terminus of striated alpha-tropomyosin in solution reveals an unusual parallel arrangement of interacting alpha-helices. Biochemistry 42:614–619

Greenfield NJ, Huang YJ, Swapna GV, Bhattacharya A, Rapp B, Singh A, Montelione GT, Hitchcock-DeGregori SE (2006) Solution NMR structure of the junction between tropomyosin molecules: implications for actin binding and regulation. J Mol Biol 364:80–96

Greenfield NJ, Kotlyanskaya L, Hitchcock-DeGregori SE (2009) Structure of the N terminus of a nonmuscle alpha-tropomyosin in complex with the C terminus: implications for actin binding. Biochemistry 48:1272–1283

Gross SR (2013) Actin binding proteins: their ups and downs in metastatic life. Cell Adhes Migr 7:199–213

Gunning P (2008) Tropomyosin. Adv Exp Med Biol 644

Gunning P, O'Neill G, Hardeman E (2008) Tropomyosin-based regulation of the actin cytoskeleton in time and space. Physiol Rev 88:1–35

Gunning PW, Hardeman EC, Lappalainen P, Mulvihill DP (2015) Tropomyosin – master regulator of actin filament function in the cytoskeleton. J Cell Sci 128:2965–2974

Hammell RL, Hitchcock-DeGregori SE (1996) Mapping the functional domains within the carboxyl terminus of alpha-tropomyosin encoded by the alternatively spliced ninth exon. J Biol Chem 271:4236–4242

Hanson J, Lowy J (1963) The sructure of F-actin and of actin filaments isolated from muscle. J Mol Biol 6:14

Helfman DM, Flynn P, Khan P, Saeed A (2008) Tropomyosin as a regulator of cancer cell transformation. Adv Exp Med Biol 644:124–131

Hendricks M, Weintraub H (1981) Tropomyosin is decreased in transformed cells. Proc Natl Acad Sci U S A 78:5633–5637

Hitchcock-DeGregori SE (2008) Tropomyosin: function follows structure. Adv Exp Med Biol 644:60–72

Hitchcock-DeGregori SE, An Y (1996) Integral repeats and a continuous coiled coil are required for binding of striated muscle tropomyosin to the regulated actin filament. J Biol Chem 271:3600–3603

Hitchcock-DeGregori SE, Song Y, Moraczewska J (2001) Importance of internal regions and the overall length of tropomyosin for actin binding and regulatory function. Biochemistry 40:2104–2112

Hodges RS, Saund AK, Chong PC, St-Pierre SA, Reid RE (1981) Synthetic model for two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization of an 86-residue analog of tropomyosin. J Biol Chem 256:1214–1224

Hodges AR, Krementsova EB, Bookwalter CS, Fagnant PM, Sladewski TE, Trybus KM (2012) Tropomyosin is essential for processive movement of a class V myosin from budding yeast. Curr Biol 22:1410–1416

Holmes KC, Lehman W (2008) Gestalt-binding of tropomyosin to actin filaments. J Muscle Res Cell Motil 29:213–219

Hsiao JY, Goins LM, Petek NA, Mullins RD (2015) Arp2/3 complex and cofilin modulate binding of tropomyosin to branched actin networks. Curr Biol 25:1573–1582

Hundt N, Steffen W, Pathan-Chhatbar S, Taft MH, Manstein DJ (2016) Load-dependent modulation of non-muscle myosin-2A function by tropomyosin 4.2. Sci Rep 6:20554

Ishii Y, Lehrer SS (1985) Fluorescence studies of the conformation of pyrene-labeled tropomyosin: effects of F-actin and myosin subfragment 1. Biochemistry 24:6631–6638

Ishii Y, Hitchcock-DeGregori S, Mabuchi K, Lehrer SS (1992) Unfolding domains of recombinant fusion alpha alpha-tropomyosin. Protein Sci 1:1319–1325

Kawai M, Lu X, Hitchcock-Degregori SE, Stanton KJ, Wandling MW (2009) Tropomyosin period 3 is essential for enhancement of isometric tension in thin filament-reconstituted bovine myocardium. J Biophys 2009:380967

Kee AJ, Hardeman EC (2008) Tropomyosins in skeletal muscle diseases. Adv Exp Biol Med 644:15

Kee AJ, Yang L, Lucas CA, Greenberg MJ, Martel N, Leong GM, Hughes WE, Cooney GJ, James DE, Ostap EM, Han W, Gunning PW, Hardeman EC (2015) An actin filament population defined by the tropomyosin Tpm3.1 regulates glucose uptake. Traffic 16:691–711

Khaitlina SY (2015) Tropomyosin as a regulator of actin dynamics. Int Rev Cell Mol Biol 318:255–291

Kirwan JP, Hodges RS (2014) Transmission of stability information through the N-domain of tropomyosin is interrupted by a stabilizing mutation (A109L) in the hydrophobic core of the stability control region (residues 97–118). J Biol Chem 289:4356–4366

Kostyukova AS (2008a) Tropomodulin/tropomyosin interactions regulate actin pointed end dynamics. Adv Exp Med Biol 644:283–292

Kostyukova AS (2008b) Tropomodulins and tropomodulin/tropomyosin interactions. Cell Mol Life Sci 65:563–569

Kostyukova AS, Rapp BA, Choy A, Greenfield NJ, Hitchcock-DeGregori SE (2005) Structural requirements of tropomodulin for tropomyosin binding and actin filament capping. Biochemistry 44:4905–4910

Kostyukova AS, Choy A, Rapp BA (2006) Tropomodulin binds two tropomyosins: a novel model for actin filament capping. Biochemistry 45:12068–12075

Kostyukova AS, Hitchcock-DeGregori SE, Greenfield NJ (2007) Molecular basis of tropomyosin binding to tropomodulin, an actin-capping protein. J Mol Biol 372:608–618

Kovar DR, Sirotkin V, Lord M (2011) Three's company: the fission yeast actin cytoskeleton. Trends Cell Biol 21:177–187

Kwok SC, Hodges RS (2003) Clustering of large hydrophobes in the hydrophobic core of two-stranded alpha-helical coiled-coils controls protein folding and stability. J Biol Chem 278:35248–35254

Laki K, Maruyama K, Kominz DR (1962) Evidence for the interaction between tropomyosin and actin. Arch Biochem Biophys 98:323–330

Landis CA, Bobkova A, Homsher E, Tobacman LS (1997) The active state of the thin filament is destabilized by an internal deletion in tropomyosin. J Biol Chem 272:14051–14056

Lehman W, Hatch V, Korman V, Rosol M, Thomas L, Maytum R, Geeves MA, Van Eyk JE, Tobacman LS, Craig R (2000) Tropomyosin and actin isoforms modulate the localization of tropomyosin strands on actin filaments. J Mol Biol 302:593–606

Lehman W, Orzechowski M, Li XE, Fischer S, Raunser S (2013) Gestalt-binding of tropomyosin on actin during thin filament activation. J Muscle Res Cell Motil 34:155–163

Lehrer SS, Geeves MA (1998) The muscle thin filament as a classical cooperative/allosteric regulatory system. J Mol Biol 277:1081–1089

Lehrer SS, Ishii Y (1988) Fluorescence properties of acrylodan-labeled tropomyosin and tropomyosin-actin: evidence for myosin subfragment 1 induced changes in geometry between tropomyosin and actin. Biochemistry 27:5899–5906

Lehrer SS, Morris EP (1982) Dual effects of tropomyosin and troponin-tropomyosin on actomyosin subfragment 1 ATPase. J Biol Chem 257:8073–8080

Lehrer SS, Morris EP (1984) Comparison of the effects of smooth and skeletal tropomyosin on skeletal actomyosin subfragment 1 ATPase. J Biol Chem 259:2070–2072

Levitsky DI, Rostkova EV, Orlov VN, Nikolaeva OP, Moiseeva LN, Teplova MV, Gusev NB (2000) Complexes of smooth muscle tropomyosin with F-actin studied by differential scanning calorimetry. Eur J Biochem 267:1869–1877

Li Y, Mui S, Brown JH, Strand J, Reshetnikova L, Tobacman LS, Cohen C (2002) The crystal structure of the C-terminal fragment of striated-muscle alpha-tropomyosin reveals a key troponin T recognition site. Proc Natl Acad Sci U S A 99:7378–7383

Li XE, Tobacman LS, Mun JY, Craig R, Fischer S, Lehman W (2011) Tropomyosin position on F-actin revealed by EM reconstruction and computational chemistry. Biophys J 100:1005–1013

Li XE, Orzechowski M, Lehman W, Fischer S (2014) Structure and flexibility of the tropomyosin overlap junction. Biochem Biophys Res Commun 446:304–308

Lu SM, Hodges RS (2004) Defining the minimum size of a hydrophobic cluster in two-stranded alpha-helical coiled-coils: effects on protein stability. Protein Sci 13:714–726

Mahadev K, Raval G, Bharadwaj S, Willingham MC, Lange EM, Vonderhaar B, Salomon D, Prasad GL (2002) Suppression of the transformed phenotype of breast cancer by tropomyosin-1. Exp Cell Res 279:40–51

Mason JM, Arndt KM (2004) Coiled coil domains: stability, specificity, and biological implications. Chembiochem 5:170–176

Matyushenko AM, Artemova NV, Shchepkin DV, Kopylova GV, Bershitsky SY, Tsaturyan AK, Sluchanko NN, Levitsky DI (2014) Structural and functional effects of two stabilizing substitutions, D137L and G126R, in the middle part of alpha-tropomyosin molecule. FEBS J 281:2004–2016

McIntosh BB, Holzbaur EL, Ostap EM (2015) Control of the initiation and termination of kinesin-1-driven transport by myosin-Ic and nonmuscle tropomyosin. Curr Biol 25:523–529

McLachlan AD, Stewart M (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J Mol Biol 98:293–304

McLachlan AD, Stewart M (1976) The 14-fold periodicity in alpha-tropomyosin and the interaction with actin. J Mol Biol 103:271–298

McLachlan AD, Stewart M, Smillie LB (1975) Sequence repeats in alpha-tropomyosin. J Mol Biol 98:281–291

Merkel L, Ikebe M, Hartshorne DJ (1989) Interaction of smooth muscle tropomyosin and smooth muscle myosin. Effect on the properties of myosin. Biochemistry 28:2215–2220

Meshcheryakov VA, Krieger I, Kostyukova AS, Samatey FA (2011) Structure of a tropomyosin N-terminal fragment at 0.98 A resolution. Acta Crystallogr Sect D: Biol Crystallogr 67:822–825

Minakata S, Maeda K, Oda N, Wakabayashi K, Nitanai Y, Maeda Y (2008) Two-crystal structures of tropomyosin C-terminal fragment 176–273: exposure of the hydrophobic core to the solvent destabilizes the tropomyosin molecule. Biophys J 95:710–719

Moore PB, Huxley HE, DeRosier DJ (1970) Three-dimensional reconstruction of F-actin, thin filaments and decorated thin filaments. J Mol Biol 50:279–295

Moore JR, Campbell SG, Lehman W (2016) Structural determinants of muscle thin filament cooperativity. Arch Biochem Biophys 594:8–17

Moraczewska J, Hitchcock-DeGregori SE (2000) Independent functions for the N- and C-termini in the overlap region of tropomyosin. Biochemistry 39:6891–6897

Mun JY, Previs MJ, Yu HY, Gulick J, Tobacman LS, Beck Previs S, Robbins J, Warshaw DM, Craig R (2014) Myosin-binding protein C displaces tropomyosin to activate cardiac thin filaments and governs their speed by an independent mechanism. Proc Natl Acad Sci U S A 111:2170–2175

Murakami K, Stewart M, Nozawa K, Tomii K, Kudou N, Igarashi N, Shirakihara Y, Wakatsuki S, Yasunaga T, Wakabayashi T (2008) Structural basis for tropomyosin overlap in thin (actin) filaments and the generation of a molecular swivel by troponin-T. Proc Natl Acad Sci U S A 105:7200–7205

Nishikawa Y, Inatomi M, Iwasaki H, Kurisu G (2015) Structural change in the Dynein Stalk region associated with two different affinities for the microtubule. Journal of molecular biology 428(9):1886–1896

Nitanai Y, Minakata S, Maeda K, Oda N, Maeda Y (2007) Crystal structures of tropomyosin: flexible coiled-coil. Adv Exp Med Biol 592:137–151

Orzechowski M, Li XE, Fischer S, Lehman W (2014) An atomic model of the tropomyosin cable on F-actin. Biophys J 107:694–699

O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254:539–544

Ostap EM (2008) Tropomyosins as discriminators of myosin function. Adv Exp Med Biol 644:273–282

Palm T, Greenfield NJ, Hitchcock-DeGregori SE (2003) Tropomyosin ends determine the stability and functionality of overlap and troponin T complexes. Biophys J 84:3181–3189

Pappas CT, Bliss KT, Zieseniss A, Gregorio CC (2011) The Nebulin family: an actin support group. Trends Cell Biol 21:29–37

Parry DAD (1975) Analysis of the primary sequence of alpha-tropomyosin from rabbit skeletal muscle. J Mol Biol 98:519–535

Parry DAD, Squire JM (1973) Structural role of tropomyosin in muscle regulation: analysis of the x-ray diffraction patterns from relaxed and contracting muscles. J Mol Biol 75:33–55

Parry DAD, Fraser RD, Squire JM (2008) Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J Struct Biol 163:258–269

Paul DM, Squire JM, Morris EP (2017) Relaxed and active thin filament structures reveal a new regulatory mechanism: varying tropomyosin shifts within a regulatory unit

Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A 37:205–211

Perry SV (2001) Vertebrate tropomyosin: distribution, properties and function. J Muscle Res Cell Motil 22:5–49

Phillips GN Jr (1986) Construction of an atomic model for tropomyosin and implications for interactions with actin. J Mol Biol 192:128–131

Phillips GN Jr, Lattman EE, Cummins P, Lee KY, Cohen C (1979) Crystal structure and molecular interactions of tropomyosin. Nature 278:413–417

Phillips GN Jr, Fillers JP, Cohen C (1980) Motions of tropomyosin. Crystal as metaphor. Biophys J 32:485–502

Phillips GN Jr, Fillers JP, Cohen C (1986) Tropomyosin crystal structure and muscle regulation. J Mol Biol 192:111–131

Potekhin SA, Privalov PL (1982) Co-operative blocks in tropomyosin. J Mol Biol 159:519–535

Prasad GL, Fuldner RA, Cooper HL (1993) Expression of transduced tropomyosin 1 cDNA suppresses neoplastic growth of cells transformed by the ras oncogene. Proc Natl Acad Sci U S A 90:7039–7043

Prasad GL, Masuelli L, Raj MH, Harindranath N (1999) Suppression of src-induced transformed phenotype by expression of tropomyosin-1. Oncogene 18:2027–2031

Pruyne D (2008) Tropomyosin function in yeast. Adv Exp Med Biol 644:168–186

Rao JN, Madasu Y, Dominguez R (2014) Mechanism of actin filament pointed-end capping by tropomodulin. Science 345:463–467

Rao JN, Rivera-Santiago R, Li XE, Lehman W, Dominguez R (2012) Structural analysis of smooth muscle tropomyosin alpha and beta isoforms. J Biol Chem 287:3165–3174

Raval GN, Bharadwaj S, Levine EA, Willingham MC, Geary RL, Kute T, Prasad GL (2003) Loss of expression of tropomyosin-1, a novel class II tumor suppressor that induces anoikis, in primary breast tumors. Oncogene 22:6194–6203

Redwood C, Robinson P (2013) Alpha-tropomyosin mutations in inherited cardiomyopathies. J Muscle Res Cell Motil 34:285–294

Robaszkiewicz K, Ostrowska Z, Marchlewicz K, Moraczewska J (2015) Tropomyosin isoforms differentially modulate the regulation of actin filament polymerization and depolymerization by cofilins. FEBS J 283(4):723–737

Schevzov G, Whittaker SP, Fath T, Lin JJ, Gunning PW (2011) Tropomyosin isoforms and reagents. BioArchitecture 1:135–164

Schmidt WM, Lehman W, Moore JR (2015) Direct observation of tropomyosin binding to actin filaments. Cytoskeleton 72:292–303

Sckolnick M, Krementsova EB, Warshaw DM, Lord MJ, Trybus KM (2016) Tropomyosin isoforms bias actin track selection by vertebrate myosin Va. Mol Biol Cell. 27:2889–2897

Singh A, Hitchcock-DeGregori SE (2003) Local destabilization of the tropomyosin coiled coil gives the molecular flexibility required for actin binding. Biochemistry 42:14114–14121

Singh A, Hitchcock-DeGregori SE (2006) Dual requirement for flexibility and specificity for binding of the coiled-coil tropomyosin to its target, actin. Structure 14:43–50

Singh A, Hitchcock-DeGregori SE (2007) Tropomyosin's periods are quasi-equivalent for actin binding but have specific regulatory functions. Biochemistry 46:14917–14927

Singh A, Hitchcock-DeGregori SE (2009) A peek into tropomyosin binding and unfolding on the actin filament. PLoS One 4:e6336

Sodek J, Hodges RS, Smillie LB, Jurasek L (1972) Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure. Proc Natl Acad Sci U S A 69:3800–3804

Sodek J, Hodges RS, Smillie LB (1978) Amino acid sequence of rabbit skeletal muscle alpha-tropomyosin. The COOH-terminal half (residues 142 to 284). J Biol Chem 253:1129–1136

Sousa DR, Stagg SM, Stroupe ME (2013) Cryo-EM structures of the actin:tropomyosin filament reveal the mechanism for the transition from C- to M-state. J Mol Biol 425:4544–4555

Stark BC, Sladewski TE, Pollard LW, Lord M (2010) Tropomyosin and myosin-II cellular levels promote actomyosin ring assembly in fission yeast. Mol Biol Cell 21:989–1000

Stehn JR, Haass NK, Bonello T, Desouza M, Kottyan G, Treutlein H, Zeng J, Nascimento PR, Sequeira VB, Butler TL, Allanson M, Fath T, Hill TA, McCluskey A, Schevzov G, Palmer SJ, Hardeman EC, Winlaw D, Reeve VE, Dixon I, Weninger W, Cripe TP, Gunning PW (2013) A novel class of anticancer compounds targets the actin cytoskeleton in tumor cells. Cancer Res 73:5169–5182

Stewart M, McLachlan AD (1975) Fourteen actin-binding sites on tropomyosin? Nature 257:331–333

Sumida JP, Wu E, Lehrer SS (2008) Conserved Asp-137 imparts flexibility to tropomyosin and affects function. J Biol Chem 283:6728–6734

Tang N, Ostap EM (2001) Motor domain-dependent localization of myo1b (myr-1). Curr Biol 11:1131–1135

Tao T, Lamkin M (1984) Crosslinking of tropomyosin to myosin subfragment-1 in reconstituted rabbit skeletal thin filaments. FEBS Lett 168:169–173

Tardiff JC (2011) Thin filament mutations: developing an integrative approach to a complex disorder. Circ Res 108:765–782

Tobacman LS (2008) Cooperative binding of tropomyosin to actin. Adv Exp Med Biol 644:85–94

Tojkander S, Gateva G, Schevzov G, Hotulainen P, Naumanen P, Martin C, Gunning PW, Lappalainen P (2011) A molecular pathway for myosin II recruitment to stress fibers. Curr Biol 21:539–550

Tripet B, Wagschal K, Lavigne P, Mant CT, Hodges RS (2000) Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". J Mol Biol 300:377–402

Tsao TC, Bailey K, Adair GS (1951) The size, shape and aggregation of tropomyosin particles. Biochem J 49:27–36

Ueno H (1984) Local structural changes in tropomyosin detected by a trypsin-probe method. Biochemistry 23:4791–4798

Umemoto S, Bengur AR, Sellers JR (1989) Effect of multiple phosphorylations of smooth muscle and cytoplasmic myosins on movement in an in vitro motility assay. J Biol Chem 264:1431–1436

Vandemoortele G, Gevaert K, Eyckerman S (2016) Proteomics in the genome engineering era. Proteomics 16:177–187

Vindin H, Gunning P (2013) Cytoskeletal tropomyosins: choreographers of actin filament functional diversity. J Muscle Res Cell Motil 34:261–274

von der Ecken J, Muller M, Lehman W, Manstein DJ, Penczek PA, Raunser S (2015) Structure of the F-actin-tropomyosin complex. Nature 519:114–117

Wagschal K, Tripet B, Hodges RS (1999a) De novo design of a model peptide sequence to examine the effects of single amino acid substitutions in the hydrophobic core on both stability and oligomerization state of coiled-coils. J Mol Biol 285:785–803

Wagschal K, Tripet B, Lavigne P, Mant C, Hodges RS (1999b) The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. Protein Sci 8:2312–2329

Wang CL, Coluccio LM (2010) New insights into the regulation of the actin cytoskeleton by tropomyosin. Int Rev Cell Mol Biol 281:91–128

Wawro B, Greenfield NJ, Wear MA, Cooper JA, Higgs HN, Hitchcock-DeGregori SE (2007) Tropomyosin regulates elongation by formin at the fast-growing end of the actin filament. Biochemistry 46:8146–8155

Wegner A (1980) The interaction of alpha, alpha- and alpha, beta-tropomyosin with actin filaments. FEBS Lett 119:245–248

Wegner A (1982) Kinetic analysis of actin assembly suggests that tropomyosin inhibits spontaneous fragmentation of actin filaments. J Mol Biol 161:217–227

Weigt C, Wegner A, Koch MH (1991) Rate and mechanism of the assembly of tropomyosin with actin filaments. Biochemistry 30:10700–10707

Whitby FG, Phillips GN Jr (2000) Crystal structure of tropomyosin at 7 Angstroms resolution. Proteins 38:49–59

Williams DL Jr, Greene LE, Eisenberg E (1984) Comparison of effects of smooth and skeletal muscle tropomyosins on interactions of actin and myosin subfragment 1. Biochemistry 23:4150–4155

Wolfenson H, Meacci G, Liu S, Stachowiak MR, Iskratsch T, Ghassemi S, Roca-Cusachs P, O'Shaughnessy B, Hone J, Sheetz MP (2016) Tropomyosin controls sarcomere-like contractions for rigidity sensing and suppressing growth on soft matrices. Nat Cell Biol 18:33–42

Wong L, Huang CH, Lee BW (2016) Shellfish and house dust mite allergies: is the link tropomyosin? Allergy, Asthma Immunol Res 8:101–106

Yamaguchi M, Ver A, Carlos A, Seidel JC (1984) Modulation of the actin-activated adenosinetriphosphatase activity of myosin by tropomyosin from vascular and gizzard smooth muscles. Biochemistry 23:774–779

Yamashiro S, Gokhin DS, Kimura S, Nowak RB, Fowler VM (2012) Tropomodulins: pointed-end capping proteins that regulate actin filament architecture in diverse cell types. Cytoskeleton 69:337–370

Zheng W, Barua B, Hitchcock-DeGregori SE (2013) Probing the flexibility of tropomyosin and its binding to filamentous actin using molecular dynamics simulations. Biophys J 105:1882–1892

Chapter 10
Titin and Nebulin in Thick and Thin Filament Length Regulation

Larissa Tskhovrebova and John Trinick

Contents

Abstract In this review we discuss the history and the current state of ideas related to the mechanism of size regulation of the thick (myosin) and thin (actin) filaments in vertebrate striated muscles. Various hypotheses have been considered during of more than half century of research, recently mostly involving titin and nebulin acting as templates or 'molecular rulers', terminating exact assembly. These two giant, single-polypeptide, filamentous proteins are bound *in situ* along the thick and thin filaments, respectively, with an almost perfect match in the respective lengths and

L. Tskhovrebova • J. Trinick (✉)
Astbury Centre, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK
e-mail: j.trinick@leeds.ac.uk

structural periodicities. However, evidence still questions the possibility that the proteins function as templates, or scaffolds, on which the thin and thick filaments could be assembled. In addition, the progress in muscle research during the last decades highlighted a number of other factors that could potentially be involved in the mechanism of length regulation: molecular chaperones that may guide folding and assembly of actin and myosin; capping proteins that can influence the rates of assembly-disassembly of the myofilaments; Ca^{2+} transients that can activate or deactivate protein interactions, etc. The entire mechanism of sarcomere assembly appears complex and highly dynamic. This mechanism is also capable of producing filaments of about the correct size without titin and nebulin. What then is the role of these proteins? Evidence points to titin and nebulin stabilizing structures of the respective filaments. This stabilizing effect, based on linear proteins of a fixed size, implies that titin and nebulin are indeed molecular rulers of the filaments. Although the proteins may not function as templates in the assembly of the filaments, they measure and stabilize exactly the same size of the functionally important for the muscles segments in each of the respective filaments.

Keywords Vertebrate striated muscle • Sarcomere structure • Filaments lengths • Molecular ruler hypothesis • Titin/Connectin • Nebulin

Abbreviations

BTS	N-benzyl-p-toluenesulphonamide
Hsp90a	Heat shock protein 90a
ITF	Intra-flagellar transport machinery
ML-7	Myosin kinase inhibitor (and inhibitor to other kinases)
MLCK	Myosin light chain kinase
MyBP-C	Myosin binding protein-C (C-protein)
RyR	Ryanodine receptor
Tmod	Tropomodulin
TRiC/CCT	T-complex protein-1 ring complex
UNC45b	Protein unc-45 homolog B (unc45 myosin chaperone B)

10.1   Introduction

Consideration of the problem of thick (myosin) and thin (actin) filament length regulation in vertebrate striated muscles started about 60 years ago, when it was realized that the strikingly regular banded appearance of these muscles reflected not only the regularity of packing, but also the remarkably uniform size of each of the filament types (Huxley 1957; Fig. 10.1). Length measurements of the thick
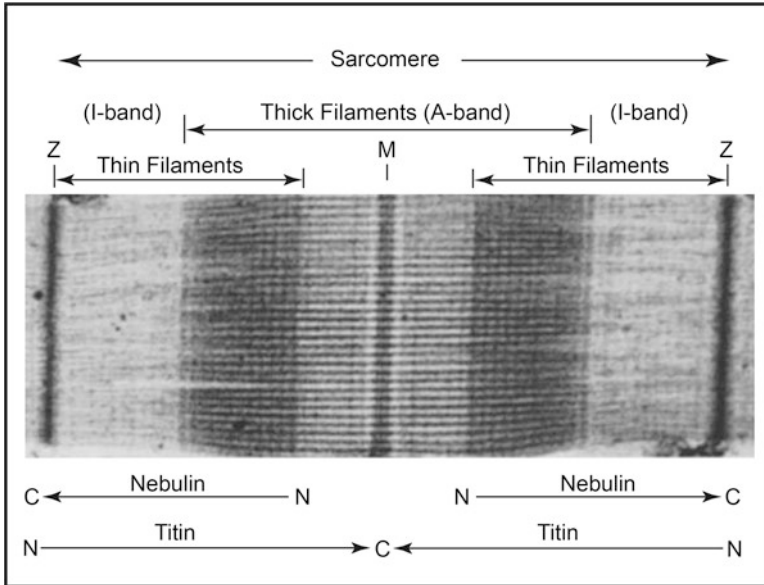
Fig. 10.1 Uniformity of the thick and thin filaments lengths in electron micrographs of muscle. Frog skeletal muscle (Adapted from Huxley HE 1967)

filaments, whether in plastic sections of muscles, negatively stained isolated A-segments, or purified filaments, usually gave values within the 1.5–1.6 μm range. This value appears to be the same in both skeletal muscles (Huxley 1963; Page and Huxley 1963; Sosnicki et al. 1991; Walker and Schrodt 1974) and in cardiac muscles (Spiro and Sonnenblick 1964; Spotnitz et al. 1966; Robinson and Winegrad 1979) of different vertebrate species. The model of the vertebrate striated muscle myosin filament, based on cryo-sections and on the observed distribution of myosin motor domains on the surface of the filaments, predicts that all vertebrate striated muscle thick filaments are 1.57 μm long and contain exactly 294 myosin molecules (Craig and Offer 1976; Sjöström and Squire 1977).

In the case of the thin filaments, size regularity is not as strict as is the case with the thick filaments. Instead of the universal length uniformity, there is a range of muscle type-related limits for thin filament length. These limits are emphasized by the relatively sharp edges of the filament arrays seen in longitudinal sections of sarcomeres. Thin filament lengths defined by these edges vary in different striated muscles in the range from ~0.9 μm to ~1.5 μm (Huxley 1963; O'Brien et al. 1971; Walker and Schrodt 1974; Kruger et al. 1991; Rinhkob et al. 2004). Further variations are also found in the lengths of the individual filaments within the arrays. It appears that up to 30 % of the filaments in an array are not reaching the maximal length limit shown by its edge (Robinson and Winegrad 1977, 1979; Traeger and Goldstein 1983; Burgoyne et al. 2008). The electron microscope data are also supported by high-resolution light microscopy (Yasuda et al. 1994; Littlefield and Fowler 2002).

Polymerization of purified myosin usually results in a Gaussian distribution of the filament lengths, with the mean value depending on the buffer composition (Harrington and Rodgers 1984). Actin polymerization in solution usually generates an exponential distribution of lengths (Kawamura and Maruyama 1970). The levels of length uniformity that are seen *in situ* are not generally reproducible *in vitro*, suggesting the presence in muscle cells of specific mechanisms regulating filament length (Huxley 1963). A vernier-type mechanism, based on two co-polymerizing protein species with different repeats, was considered (Huxley 1963; Huxley and Brown 1967). However, this idea was soon rejected since, as the following studies showed, such a mechanism could not be based either on isoforms of major proteins of the thin and thick filaments, or on their interactions with the co-polymerizing proteins. The filaments, at least in vertebrate muscles, appear to be formed mainly by single isoforms of actin (Tondeleir et al. 2009) and myosin (Eddinger 1998). The *in vitro* co-polymerization with tropomyosin (for actin filaments, Kawamura and Maruyama, 1970; Ishiwata and Funatsu 1985) or myosin binding protein-C (MyBP-C; for myosin filaments, Moos et al. 1975; Miyahara and Noda 1980; Davis 1988) did not decrease the length heterogeneity of the filaments.

A number of factors, however, have been found to affect the filament length distributions. In the case of actin filaments, polymerization in the presence of the barbed end capping proteins was shown to produce Gaussian, rather than exponential, distributions (Kuhlman 2005). In the case of myosin filaments, the effects of the specific cellular environment and the significance of an optimal combination of ion composition, ionic strength, pH and ATP in the cell were tested (Huxley 1963; Kaminer and Bell 1966; Josephs and Harrington 1966; Harrison et al. 1971; Katsura and Noda 1971; Reisler et al. 1980; Pinset-Härström 1985). All these parameters produced effects, and synthetic thick filaments with relatively narrow length distributions were obtained, although not as sharply defined as in muscle cells (Josephs and Harrington 1966; Harrington and Rodgers 1984; Davis 1988). Attention was also focused on the role of the filament length-dependent differences in the myosin-myosin association and dissociation rate constants (Davis 1981, 1985; Higuchi and Ishiwata 1985), which were reflected also in a decrease of stability with an increase in length of the filament (Trinick and Cooper 1980; Ishiwata 1981; Ishiwata et al. 1985). The phenomenon seemed to indicate the accumulation of stress in the elongating filament (Higuchi et al. 1986), a possibility previously discussed by Pepe (1967). Estimates showed, however, a general insufficiency of the stress-based mechanism to ensure the sharpness of the length distribution equivalent to that in the native filaments (Higuchi et al. 1986), suggesting a need to search for other mechanisms.

10.2 Titin and Nebulin – The "Molecular Ruler" Hypotheses

While all these mechanisms were being debated, two new giant filamentous proteins were discovered in vertebrate sarcomeres; titin, also known as connectin (Maruyama et al. 1976; Wang et al. 1979), and nebulin (Wang and Williamson

1980). Analysis of the cDNA sequences of these proteins showed that each includes a family of isoforms, differing in the sequence lengths and the respective molecular weights (Bang et al. 2001; Kazmierski et al. 2003).

10.2.1 Titin

Predicted human titin isoforms range from ~3 to ~4 MDa, with the respective sequence lengths ranging from ~27,000 to ~38,000 aa (UniProtKB – Q8WZ42 (Titin_Human); Bang et al. 2001; Li et al. 2012). The isoform size was found to correlate with the passive tension of different muscle types (Wang et al. 1991; Trombitás et al. 2000; Anderson and Granzier, 2012). The purified titin molecule is about 1 μm long and 4 nm wide (Wang et al. 1984; Nave et al. 1989; Sonoda et al. 1990), with the domain substructure reminiscent of "beads on a string" (Trinick et al. 1984). It also appears highly flexible and tends to adopt a compact coil conformation (Trinick et al. 1984; Wang et al. 1984; Tskhovrebova and Trinick 2001). *In situ* the titin molecule spans between the Z-disc (N-terminus) and the M-line (C-terminus; Fürst et al. 1988, 1989a; Whiting et al. 1989). The A-band part is bound to the backbone of the thick filament (Fürst et al. 1988; Pierobon-Bormioli et al. 1989; Whiting et al. 1989). There are six molecules of titin bound to each half of the filament (Cazorla et al. 2000; Liversage et al. 2001). Electron microscope and X-ray diffraction studies agree with the location of titin on the surface of thick filament backbones and suggest binding of three side-by-side titin dimers to each half of the filament (Cantino et al. 2002; Squire et al. 2004; Zoghbi et al. 2008; AL-Khayat et al. 2013). Biochemical and ultrastructural studies also favour A-band titin occurring as dimers *in situ* (Tskhovrebova et al. 2010; AL-Khayat et al. 2013).

The I-band part of titin forms a flexible connection between the tip of the thick filament and the Z-disc. In agreement with its *in situ* location, the native protein (Fürst et al. 1992; Soteriou et al. 1993) and recombinant fragments of A-band titin (Labeit et al. 1992) show a binding affinity to myosin and to MyBP-C. The beaded substructure reflects the modular domain organization consisting mostly of about 300 Ig and Fn3 domains (Labeit et al. 1990; Bang et al. 2001). In A-band titin, the Ig and Fn3 domains are arranged into "super-repeats". The size of the super-repeats and their internal subdivisions imply periodicities of ~14 nm and ~43 nm, which closely match the repeats of the myosin and MyBP-C arrangements in the thick filament (Fig. 10.2). The high level of inter-species sequence conservation here also agrees with the apparent conservation of thick filament structure.

10.2.2 Nebulin

Nebulins are smaller than titins; their sequence lengths being between 6669 aa (~773 kDa) and 8525 aa (~1 MDa) (UniProtKB – P20929 (Nebu-Human); Labeit and Kolmerer 1996; Wang et al. 1996; Zhang et al. 1996; Donner et al. 2004;
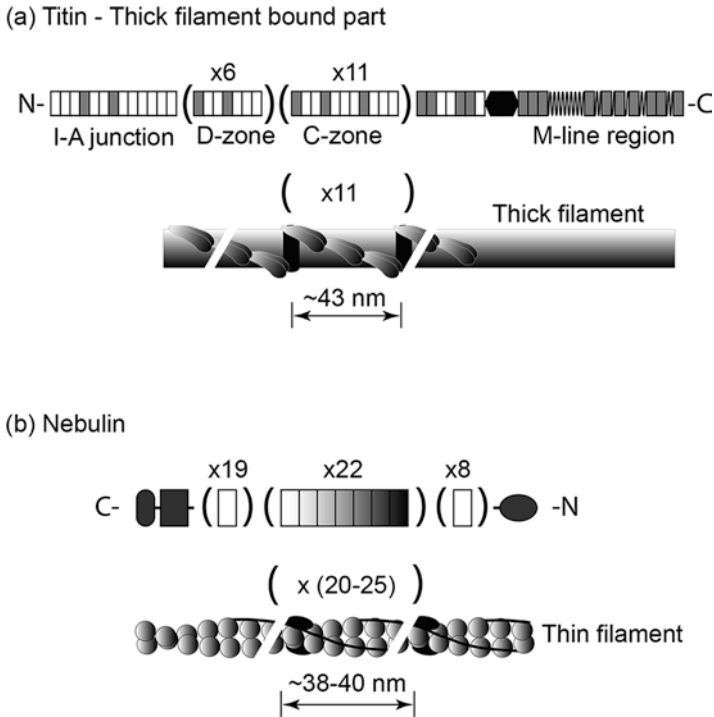
Fig. 10.2 Correlation between the super-repeat organization of titin (**a**) and nebulin (**b**) molecules and the structural periodicities of the thick and thin filaments, respectively. (**a**) Most of the thick filament bound A-band part of titin is formed by immunoglobulin (in *grey*) and fibronectin (in *white*) domains, each containing ~100 amino acid residues. The domains are arranged into patterns forming super-repeats. The patterns suggest ~14 nm periodicity. There are six 7-domain super-repeats (the D-zone region, corresponding to the tapered distal part of the thick filament) and eleven 11-domain super-repeats (the C-zone region, corresponding to the central most uniform part in each half of the thick filament, containing also MyBP-C spaced at ~43 nm). The size of the super-repeats in the D-zone region is ~28 nm, and in the C-zone region ~43 nm. In *black* is shown the kinase domain, and in zigzag, the unique PEVK sequences. (**b**) Most of the nebulin molecule is also formed by sequence repeats (in *white*), each of ~ 35 residues, or ~ 5 nm long, corresponding approximately to the 5.5 nm actin monomer separation in a long period strand of an actin filament. The N- and C-terminal regions contain small groups of simple repeats. The central >150 repeats are arranged into super-repeats, each containing seven simple repeats. The super-repeats predict periodicity of about 38 nm, matching to the helical and tropomyosin-troponin periodicities of the thin filament

Hanashima et al. 2009). They are integral components of native thin filaments (Wang and Wright 1988; Maruyama et al. 1989; Pierobon-Bormioli et al. 1989). The molecule spans the entire length of the thin filament with the C-terminus at the Z-disc and the N-terminus at the distal pointed end of the filament (Wright et al. 1993). Differences in nebulin isoform sizes appear to correlate with the differences in thin filament lengths in different muscles (Kruger et al. 1991; Labeit and Kolmerer 1995). A few attempts made so far to purify nebulin suggest that the resting

conformation is also a compact coil (Pierobon-Bormioli 1989; Yadavalli et al. 2009; Chitose et al. 2010). Purified nebulin (Chitose et al. 2010), as well as its recombinant fragments (Jin and Wang 1991; Pfuhl et al. 1994; Shih et al. 1997) interact *in vitro* with actin, in agreement with the *in situ* data. cDNA analysis predicts that more than 90 % of the polypeptide consists, as in the titin case, of homologous sequence repeats arranged further into super-repeats (Labeit et al. 1991; Labeit and Kolmerer 1995; Zhang et al. 1996; Wang et al. 1996). The repeats are likely to have an alpha-helical fold, although the formation of a coiled coil structure is unlikely (Labeit et al. 1991; Chen and Wang 1994; Pfuhl et al. 1994). With the absence of a known tertiary structure, the length of the alpha-helical isoform of nebulin could be estimated to be of about the same as that of the thin filament in the respective muscle. Sequence repeats suggest periodicities of about 5.5 and 38.5 nm along the nebulin helix, which match periodicities of actin subunits and the TM-TN regulatory complex in the thin filament (see Chaps. 9 and 11). This correspondence suggests multiple periodically spaced interactions between nebulin and the actin filament.

10.2.3   The Ruler Hypotheses

With both giant proteins, a close correspondence is found between their sizes and the molecular architectures and sizes of the polymer filaments to which they are bound (Fig. 10.2). Both have multiple regularly spaced monomer-binding sites with intervals matching the structural periodicities of the polymers. This close correspondence prompted the hypotheses that the proteins function as molecular rulers/templates defining the lengths of the myofilaments during muscle assembly. So A-band titin would control thick filament length (Whiting et al. 1989), and nebulin would define thin filament length (Jin and Wang 1991; Kruger et al. 1991; Labeit et al. 1991; Trinick 1992; Wright et al. 1993). Since then a large volume of information related to the role of the proteins in length regulation of the thin and thick filaments has been accumulated. One of the major experimental approaches that tested the hypotheses was analysis of the sequence of events during muscle development (reviewed by Fürst and Gautel 1995; Kontrogianni-Konstantopoulos et al. 2009; Sanger et al. 2010; Yang et al. 2014). Various animal models and gene-interference based methods have also been employed. Below is given a very short overview of some of these studies.

10.3   Titin/Nebulin in Thick/Thin Filament Assembly During Sarcomerogenesis

Studies of myofibrillogenesis have been based on both electron and immunofluorescent microscopy. In both cases, the process was described as highly dynamic and not synchronous across the cell. There is also some variability and divergence in the observations, probably reflecting the wide range of sample types used in the studies. Most of the electron microscopy was done before titin and nebulin were discovered,

and mainly described events related to the appearance and integration of the thin and thick filaments into sarcomere structure. Immunofluorescent microscopy, in turn, provided important information on the sequence of appearance of most, if not all, sarcomere proteins, as well as on the sequence of their alignment within the periodic sarcomere pattern over time.

10.3.1 Electron Microscopy

The major structural patterns that are usually seen during sarcomerogenesis are:

1. Thin filament networks and individual thin and thick filaments
2. Loose bundles of the thin filaments with patches of optically denser material, the Z-bodies, containing alpha-actinin (Lazarides and Burridge 1975). The bundles, or I-Z-I complexes, are occasionally interdigitated with small arrays of parallel thick filaments, but not in perfect register and individual thick filaments may also be present
3. Completed mature sarcomere arrangements with interdigitated arrays of thin and thick filaments in perfect register and the Z-bodies narrowed to the thinner Z-discs (Allen and Pepe 1965; Obinata et al. 1966; Kelly 1969; Myklebust et al. 1978; Sœtersdal et al. 1980; Hay 1963; Dessouky and Hibbs 1965; Fischman 1967; Shimada et al. 1967)

In some cases, the actin and myosin filaments are seen to appear in different cellular locations, emphasizing their independent assembly. Actin filaments appear mostly in proximity the cellular membrane or other membranous structures (smooth-surfaced SR tubules), whereas myosin filaments, at least in some studies, are mainly seen in the central, ribosome-rich areas (Kelly 1969; Myklebust et al. 1978). Some, but not all, studies report also that the first thick filaments, possibly the thin filaments as well, have lengths equal to, or at least not exceeding those in the adult muscle, i.e., 1.5–1.6 μm, and ~1 μm, respectively (Allen and Pepe 1965; Obinata et al. 1966; Fischman 1967; Hagopian and Spiro 1968).

10.3.2 Immunofluorescent Microscopy

The consensus sequence of events during myogenesis is the following: (1) Titin is one of the first sarcomeric proteins expressed at the start of sarcomerogenesis (Hill et al. 1986; Tokuyasu and Maher 1987a; Wang et al. 1988; Fürst et al. 1989b; Schaart et al. 1989; Colley et al. 1990; van der Loop et al. 1996a, b; Begum et al. 1998), whereas nebulin is one of the last (Fürst et al. 1989b; Moncman and Wang 1996; Begum et al. 1998). (2) Soon after expression, both titin (Tokuyasu and Maher 1987b; Fürst et al. 1989b; Schultheiss et al. 1990; Lin et al. 1994; Rhee et al. 1994; Ehler et al. 1999; Rudy et al. 2001), and nebulin (Komiyama et al. 1992; Lin et al.

1994), become anchored to I-Z-I complexes in the regions of the Z-bodies, co-localizing there with α-actinin and actin (Tokuyasu and Maher 1987b; Fürst et al. 1989b; Schultheiss et al. 1990; Komiyama et al. 1992; Lin et al. 1994; Ehler et al. 1999; Rudy et al. 2001). (3) The early anchoring of titin to the Z-bodies, involves only the N-terminal end of the molecule (Fürst et al. 1989b; Schultheiss et al. 1990; Komiyama et al. 1993; Lin et al. 1994), and the C-terminal/A-band part is predicted to be in a coiled state (Van der Loop et al. 1996b), even if it is bound to myosin (Isaacs et al. 1992). (4) The early anchoring of nebulin to the Z-bodies is likely to involve only the C-terminal part, leaving the rest either free floating or only loosely bound to the actin filaments (Begum et al. 1998). (5) Alignment of A-band titin into its final sarcomere pattern is seen by immuno-fluorescence to either precede, or occur simultaneously with, the alignment of myosin into characteristic adult sarcomeric A-bands of ~1.6 μm width, and in parallel with the alignment of the accessory (MyBP-C) and M-line (myomesin) proteins (Wang et al. 1988; Schultheiss et al. 1990; Handel et al. 1991; van der Loop et al. 1992; Ehler et al. 1999; van der Ven et al. 1999). (6) Alignment of actin-nebulin pairs into the striated pattern of the sarcomere appears to occur after that of myosin-titin (Wang et al. 1988; Nwe et al. 1999; Tokuyasu and Maher 1987a; Wang et al. 1988). However, in respect to the relative order of alignment of actin and nebulin, the observations seem to diverge (Moncman and Wang 1996; Begum et al. 1998). In the process of alignment, nebulin was seen to re-arrange, gradually extending along the thin filaments on both sides of the Z-bodies until it reached the free ends of the filaments (Begum et al. 1998).

Immuno-fluorescent studies led to three major models of sarcomerogenesis (Sanger et al. 2005, 2010): (1) the "template model", suggesting the stress-fiber-like structures as templates for sarcomere/myofibril assembly (Dlugosz et al. 1984), (2) the "independent assembly" model, suggesting that arrays of actin filaments with Z-bodies/Z-discs and arrays of myosin filaments are assembled independently and subsequently integrate into sarcomeres/myofibrils, with titin acting as an integrator (Schultheiss et al. 1990; Lu et al. 1992; Holtzer et al. 1997), and (3) the "pre-myofibril" model, suggesting that myofibril assembly involves gradual elongation of the sarcomeres from a mini-size to the mature length. During this process preformed sarcomeric myosin filaments of 1.6 μm length would incorporate into the mini-sarcomere structure displacing the short non-muscle myosin filaments (Rhee et al. 1994; Sanger et al. 2010).

10.3.3   Relevance to the Titin/Nebulin Ruler-Template Hypotheses

(1) Titin expression and alignment into the striated sarcomere pattern in advance or simultaneously with myosin agrees with the titin-template role (Ehler et al. 1999; van der Ven et al. 1999). (2) Incorporation of pre-formed full-length 1.6 μm long

sarcomeric myosin filaments into actin filaments arrays, suggested by the sarcomerogenesis models, does not agree with the titin-template hypothesis, and suggests instead a titin-integrator role (Schultheiss et al. 1990; Holtzer et al. 1997). (3) Incorporation of pre-formed half-length long (~0.8 μm) sarcomeric myosin filaments into actin filaments arrays and their growth to the full length within the sarcomere does agree with the titin-ruler role (Du et al. 2008a). (4) Nebulin expression and alignment into the striated sarcomere pattern after actin polymerization does not agree with the nebulin-template role (Begum et al. 1998).

It should also be noted that these studies did not distinguish between the adult and embryonic isoforms of titin and nebulin, or between different adult isoforms of myosin. Embryonic isoforms of both titin (Lahmers et al. 2004; Opitz et al. 2004; Warren et al. 2004; Ottenheijm et al. 2009; Li et al. 2012) and nebulin (Wang et al. 1996; Buck et al. 2010), as well as cardiac muscle specific nebulette (a mininebulin; Millevoi et al. 1998; Arimura et al. 2000; Ogut et al. 2003; Esham et al. 2007), differ from the adult isoforms. The embryonic titin isoforms, in both skeletal and cardiac muscles, are longer than the adult isoforms. The extra length is mainly due to a longer I-band part, which includes extensions of both the unique regions (PEVK) and the Ig-tandem segments. In the nebulin case, developmental changes seem to involve mainly the C-terminal end and the N-terminal super-repeats of the molecule (Donner et al. 2004, 2006; Buck et al. 2010). Developmental changes in titin isoforms indicate changes in the elastic and possibly signaling properties of the protein (Warren et al. 2004; Guo et al. 2010), whereas changes in nebulin/nebulette isoforms are likely to be associated with restructuring of the Z-bodies into Z-discs. How the isoform changes correlate with the different stages of sarcomere assembly has to be established. There is also a probability of further isoform changes and the presence of more than one isoform (Trombitás et al. 2001; Seeley et al. 2007).

10.4   Thick Filament Assembly with Absent/Truncated Titin/Nebulin

10.4.1   *Titin Truncations*

In the case of titin, impacts on sarcomerogenesis of several different types of gene mutation, leading to expression of truncated proteins, have been analyzed. In all these cases sarcomerogenesis was disrupted. In the cases when titin translation was inhibited (Person et al. 2000), or only the Z-disc segment of titin was expressed while the rest was truncated (van der Ven et al. 2000), the observed failure of myosin to align into the A-bands was in favour of the ruler-template hypothesis. However, in later studies involving longer expressed segments of the truncated protein (Xu et al. 2002; Gotthardt et al. 2003; Miller et al. 2003; Peng et al. 2005; Musa et al. 2006; Weinert et al. 2006; Peng et al. 2007; Seeley et al. 2007; Gramlich et al. 2009; Myhre et al. 2014) it was found that the initial stages of sarcomere assembly were

unaffected; sarcomere assembly did occur without titin (Xu et al. 2002; Weinert et al. 2006; Seeley et al. 2007; Myhre et al. 2014). Importantly, even in the cases when the entire A-band part of titin, normally bound to the thick filament, was absent, the size of the assembled A-bands appeared the same as in normal muscles, suggesting that the absence of full-length titin did not eliminate the thick filament length-control mechanism. However, the assembled sarcomeres appeared to lack the structural stability of the wild-type sarcomeres and, as a result, the start of contractile activity led to their disintegration (Myhre et al. 2014). These results directly contradict the ruler-template role of titin, suggesting instead a role in sarcomere organization and stability of the A-bands and thick filaments.

10.4.2   Nebulin Truncations

Experiments testing the nebulin-ruler hypothesis have, as with titin, also produced contradictory results. Thus, several groups presented evidence suggesting that the absence of nebulin in muscle cells disrupts the thin filament length regulation mechanism, favouring the ruler-template hypothesis (McElhinny et al. 2005; Bang et al. 2006; Witt et al. 2006). The observed effects, however, differed between the studies and seemed to indicate dependence on the method used. Thus, inhibition of nebulin translation in rat cardiomyocytes had an elongating effect on the thin filaments (McElhinny et al. 2005), whereas mouse models with a knockdown nebulin gene illustrated a shortening effect on the thin filaments (Bang et al. 2006; Witt et al. 2006). It was also found that muscle usage in mutants causes progressive misalignment of myofibrils and degeneration of Z-discs, and eventually leads to the disintegration of sarcomere structure (Bang et al. 2006).

Other experiments directly contradicted the ruler-template hypothesis. Thus, replacement of endogenous nebulin in chick skeletal myocytes with a shorter synthetic mini-nebulin (~0.2 μm *vs* ~1 μm), which was expected to prevent growth of the filaments beyond a 0.2 μm size, did not prevent this, and the resulting filaments grew to the about the same length as in controls (Pappas et al. 2010). At the same time, the structural stability of the filaments beyond the mini-nebulin length was found to be significantly lower than in the region bound to mini-nebulin (Pappas et al. 2010), emphasizing a stabilizing effect of nebulin on the thin filament structure, in agreement with earlier predictions (Chen et al. 1993).

Myofibril misalignment and anomalous widening and fragmentation of the Z-discs appear to be the most common observations in the nebulin-deficient skeletal muscles (Bang et al. 2006; Witt et al. 2006; Tonino et al. 2010). Similar progressive disorders in the Z-discs were also observed in cardiac muscles of mice models with knocked down nebulette (Mastrototaro et al. 2015), a cardiac muscle specific mini-isoform of nebulin (Moncman and Wang 1995). Overall, the observations indicate the importance of nebulin for the maintenance of integrity and connectivity of myofibrils and for their mechanical stability during contractile function, specifically for

the structural stability of the thin filaments and the force bearing and transmitting Z-discs.

10.5 Thick and Thin Filament Length Control – Possible Involved Factors

In summary of the above discussion, the current evidence is rather against the proposed titin/nebulin template functions, while emphasizing their structure-stabilizing roles in the thick and thin filaments, respectively. At the same time, evidence was accumulated regarding the involvement in filament length regulation of other muscle proteins and of non-protein factors.

10.5.1 Tropomodulin

One protein with a proposed role in length regulation of thin filaments is tropomodulin (Tmod), a globular actin-capping protein of MW ~ 42 kDa, binding at the pointed ends of the filaments (reviewed by Dos Remedios et al. 2003; Yamashiro et al. 2012). Tmod is shown to have binding affinity to all three major components of the filament, i.e., to actin, tropomyosin (Weber et al. 1994), and to the N-terminal end of nebulin located close to the pointed end of the thin filament (McElhinny et al. 2001). Earlier, it was thought that the capping function of Tmod is coupled with the length-measuring function of nebulin (McElhinny et al. 2005; Fowler et al. 2006). It appears now that this may not be the case and Tmod may function independently. Immunofluorescent studies of the relative positions of the C-terminus of nebulin and the actual pointed ends of the thin filaments capped by Tmod led to discovery that nebulin does not actually extend to the tip of the filament, as was expected, leaving a nebulin-free end on average ~ 15 % of the filament length (Castillo et al. 2009; Gokhin et al. 2010; Greaser and Pleitner 2014). The apparent difference in length between the actin polymer and nebulin casts doubt on nebulin's role as a thin filament template-ruler. At the same time there are findings that the inhibition of the interaction between the thin filaments and Tmod (Gregorio et al. 1995), or the knocking down of Tmod (Gokhin et al. 2015) interferes with the length control. These observations appear to favour Tmod as one of the key players in the mechanism.

10.5.2 N-RAP

N-RAP, a multidomain filamentous protein, closely related to nebulin, but about a fifth of its size (reviewed by Crawford and Horowits 2011; Bang and Chen 2015; Chu et al. 2016), is suggested to form a scaffold for I-Z-I complexes at the start of

myofibril assembly and then to participate in formation of new sarcomeres at the ends of myofibrils in the mature muscles (Carroll et al. 2004; Manisastry et al. 2009). The conclusion is based on three major lines of observations. Firstly, N-RAP has a binding affinity to the Z-bodies/Z-discs proteins, actin and α-actinin, and to the proteins linking actin filaments to the cell membrane, filamin, vinculin and talin (Luo et al. 1999; Lu et al. 2003). Secondly, during myofibrillogenesis, N-RAP appears in the cell simultaneously, or before α-actinin (and thus before nebulin), and assembles in fibrillar structures, which appear to provide binding sites to α-actinin/ Z-bodies (Manisastry et al. 2009). Thirdly, with maturation of the sarcomere structure, N-RAP moves, concentrating mainly at the myofibril ends, linking the thin filaments to the cell membrane, i.e., at the sites where new sarcomeres are likely to form (Carroll and Horowits 2000). It was proposed that N-RAP controls the antiparallel arrangement of the thin filaments in the Z-bodies. These observations give an apparent priority to N-RAP as a possible scaffolding/ruling protein in comparison to nebulin.

10.5.3   Chaperones

It has long been known that both prokaryotic and eukaryotic cells respond to the temperature shock or other type of stress by expressing special sets of proteins, the 'heat-shock proteins' (Hsp) (Alahiotis 1983; Pelham 1985). The mode of action of these is thought to include association with the stress-damaged/unfolded polypeptides, and either mediation of their folding, in which case Hsp would function as 'molecular chaperones', or transmitting the substrates to the protein-degradation systems (reviewed by Craig et al. 1993; Hendrick and Hartl 1993; Parsell and Lindquist 1993). Eukaryotes have two different sets of chaperones, one of which functions as a general stress-related system, and the other being specifically coupled with the polypeptide translation machinery, mediating folding of the newly synthesized polypeptides (reviewed by Frydman 2001; Albanèse et al. 2006; Richter et al. 2010).

    In striated muscles, members of several chaperone families have been implicated in mediating folding and assembly of myosin heavy chains: Hsp90a (heat-shock protein, ~90 kDa), UNC-45b (UCS-domain containing protein), and, at least two members of the SMYD family (SET and MYND domain-containing proteins), SMYD1 and SMYD2 (reviewed by Myhre and Pilgrim 2012; Du et al. 2014; Smith et al. 2014). The importance of the chaperones for muscle development and function is illustrated by the disruptive consequences of their deficiency (Tan et al. 2006; Wohlgemuth et al. 2007; Etard et al. 2007; Du et al. 2008b; Geach and Zimmerman 2010; Just et al. 2011; Donlin et al. 2012; Voelkel et al. 2013; Nagandla et al. 2016). Hsp90a and UNC45b (Etard et al. 2008), as well as SMYD1 (Just et al. 2011), are shown to co-localize with myosin during sarcomere assembly. SMYD2 localizes predominantly in the cytoplasm, and is shown to methylate Hsp90a and to form triple complexes with the skeletal titin isoform by binding at its unique N2A region (Donlin et al. 2012).

The chaperones are capable of interacting not only with myosin but also with each other (Barral et al. 2002; Etard et al. 2007; Srikakulam et al. 2008; Donlin et al. 2012; Li et al. 2013), suggesting a possibility of functional cooperation in the complexes. In addition, UNC-45 has been shown to self-associate forming chain-like oligomers (Gazda et al. 2013). This type of oligomerization, with the apparent match between the subunit periodicity in the chaperone chain and the structural periodicity of the thick filaments, suggests a possibility for simultaneous control of myosin folding and assembly. Folding of several nascent myosin chains could be processed concurrently by the subunits of a chain-like chaperone complex, so that the folded myosin molecules may bind to each other without being released into the cytoplasm (Gazda et al. 2013; Pokrzywa and Hoppe 2013).

Information is more limited regarding involvement of chaperones in the folding/ assembly of sarcomeric actin. However, numerous studies implicate chaperonin TRiC (T-complex protein-1 ring complex, also known as CCT), an oligomeric protein with molecular weight ~900 kDa, in mediation of folding of the actin isoforms in eukaryotes (Sternlicht et al. 1993; Eggers et al. 1997; Thulasiraman et al. 1999; reviewed by Dunn et al. 2001; Spiess et al. 2004). The possibility that folding of the alpha-actin isoform specific for vertebrate striated muscle can also be mediated by the chaperonin was also illustrated (e.g., Llorca et al. 1999; Altschuler et al. 2005). It is thought that the chaperonin mediates folding but not polymerization of actin (Dunn et al. 2001). However, it has also been noted that the majority of the substrates of this chaperonin are the proteins that form homo- or oligomeric complexes (Spiess et al. 2004). An example with tubulin, one of the substrates of the chaperonin, shows that, after completion of folding, the tubulin monomer is transferred to other chaperones which mediate its assembly into filaments (Dunn et al. 2001).

Involvement of chaperones in the folding and assembly of actin and myosin would probably not compromise the titin/nebulin ruler models, but would add complexity to the mechanism, especially as titin and nebulin themselves are also likely to be substrates for the chaperones (Golenhofen et al. 2002; Bullard et al. 2004; Voelkel et al. 2013).

10.5.4   Contractile Activity

Spontaneous changes in the intracellular Ca^{2+} concentration and the early contractile cycles, as well as interactions with the extracellular matrix are known to be important regulatory factors of sarcomerogenesis in developing muscle cells (De Deyne 2000; reviewed by Samarel 2005; Ferrari et al. 2006; Sparrow and Schock 2009; Tu et al. 2016). As an example, three different types of spontaneous and repetitive Ca^{2+} transients, characterized by distinct spatiotemporal patterns, were observed in the C2C12 mouse muscle cell line during myogenesis (Lorenzon et al. 1997). Different types of Ca^{2+} transients that occur prior to sarcomere assembly and before development of electrical excitability and contractile activity in developing muscle cells were also observed in *Xenopus* embryos (Ferrari et al. 2006). The

occurrence of the transients was attributed to the spontaneous and Ca^{2+}- and/or acetylcholine-induced openings of the sarcoplasmic reticulum Ca^{2+} release channels, the ryanodine receptors (RyR). It was suggested that, as in other cell types (Spitzer 1994; Gu and Spitzer 1995), the Ca^{2+} transients might encode in their frequency and amplitude a development program for muscle assembly. Inhibition of Ca^{2+} transients in *Xenopus* embryos by chemical blocking of RyR receptors was shown to interfere with somite maturation and myofibrillar organization (Ferrari and Spitzer 1999), whereas mutant zebra fish embryos, lacking acetylcholine receptors, assembled sarcomeres that were ~10 % shorter than in control animals (Brennan et al. 2005).

The major effect caused by interference with the acto-myosin interactions, as observed by immunofluorescence, is related to inhibition of striated sarcomere pattern formation. Such disorders have been illustrated, e.g., in the studies of chicken skeletal muscle myocytes (Soeno et al. 1999). The actin-myosin interactions in these experiments were inhibited with BDM (2,3-butanedione 2-monoxime), an inhibitor of the myosin ATPase. In contrast to the control cells, immunofluorescence showed that the striated pattern related to the mature sarcomere positions of the contractile proteins was not formed. Electron microscopy also confirmed that, although the thick and thin filaments were mostly arranged in parallel, the arrangement was irregular, and, instead of thin Z-discs periodically spaced along myofibrils, irregular distributions of Z-bodies were observed within thin filament bundles. Similar effects with sarcomere/myofilament disorganization on the same species were also observed after application of BTS (N-benzyl-p-toluenesulphonamide), a more specific inhibitor of skeletal muscle myosin-II ATPase (Kagawa et al. 2006). Analogous studies on myocyte cultures of *Xenopus* embryos also confirmed disruptive effects on myofibrillogenesis related to inhibition of contractile activity (Ramachandran et al. 2003). Recent similar experiments indicated that BTS-treatment delays lateral alignment of Z-bodies and maturation of Z-discs (Geach et al. 2015), corroborating observations of Soeno et al. (1999). The conclusion is consistent with the proposal on the basis of modelling of a role for mechanical tension in speeding up assembly and lateral registration of nascent myofibrils (Friedrich et al. 2011).

Interesting observations were also made on the early actin-titin-myosin relationships by studying the effects of Ca^{2+} transients and kinase activity on myogenesis in *Xenopus* embryos (Harris et al. 2005). Blocking Ca^{2+} transients by ryanodine early in myogenesis disrupted assembly of the proteins into sarcomeres. At the same time, myosin-titin association appeared to be enhanced, while actin-titin association weakened, as could be judged by the observed co-localization of respective pairs of the proteins in the cells. However, blocking Ca^{2+} transients later in myogenesis did not cause significant effects, at least in the case of titin. These results indicated that, while the titin-myosin interaction does not depend, or depends only weakly on, Ca^{2+} transients, titin-myosin alignment into normal sarcomere structure strongly depends on it. Furthermore, application of ML-7 (the myosin light chain kinase (MLCK) and other kinases inhibitor) during myogenesis was found to cause a correlated disorganization of titin and myosin without affecting actin alignment into striated pattern. This correlated disruption indicated an involvement of either MLCK or titin kinase, or both into the last stages of titin and myosin assembly into sarcomere structure.

The mechanism relating contractile activity to the 'maturation' of sarcomere structure and alignment of the filaments into periodic pattern is not known. It was noted, however, that the start of contractile activity does not require completion of sarcomere assembly (van der Loop et al. 1992). From a general point of view, it would be expected that the active tension developed by the first acto-myosin interactions not only would assist parallel alignment and registration of the filaments in nascent myofibrils, as discussed above, and could also cause re-organization of the force-bearing protein networks, such as Z- and M-regions, in nascent contractile assemblies, unless their passive tension is not balancing or exceeds the active force. In respect to the filament lengths, it could also be hypothesized that the contraction-caused collision of the thin and thick filament arrays with the dense protein networks of the M- and Z-regions, respectively, could destabilize subunits at the tips of the longest filaments, resulting in 'trimming' the bundles by mechanical force.

10.6 Thick and Thin Filaments Length Control in Striated Muscles of Invertebrates (Indirect Flight Muscles of *Drosophila*)

Important observations, concerning filament length regulation in striated muscles, were made in studies of sarcomerogenesis in the indirect flight muscles of *Drosophila*. It must be noted that, although filaments lengths in these muscles also appear uniform, titin and nebulin molecules spanning the entire filament lengths are not present. Instead, as all striated muscles of invertebrates, these muscles contain shorter members of the titin and nebulin protein families (Bullard et al. 2002, 2005; Tskhovrebova and Trinick 2003; Pappas et al. 2011; Bang and Chen 2015). Therefore no 'molecular ruler' mechanism, analogous to that proposed for vertebrates could function in invertebrates. Indirect flight muscles of *Drosophila*, for instance, contain the A-band titin-like protein projectin and several splice products of the *sallimus* gene (*Drosophila* titins), which have I-band titin-like structures and function as elastic connectors between the tip of the thick filament and the Z-disc (Bullard et al. 2002, 2005). There is also a nebulin-like protein *Lasp-1*, which has only two nebulin-like repeats and is localized in the Z-disc regions (Fernandes and Schöck 2014).

Myofibrillogenesis in indirect flight muscles of *Drosophila* (Reedy and Beall 1993) appears to follow, at least in major aspects, the 'pre-myofibril' model, which is currently the most popular model of myofibrillogenesis in vertebrates (Sanger et al. 2005, 2010). Briefly, this starts from formation of myofibril-like assemblies, in which the 'sarcomeres', and the thick and thin filaments too, are much shorter compared to those in mature muscles. The width of the nascent myofibrils is also small, containing just a few myofilaments in cross-sections. During transition to the adult myofibril pattern, an increase in length of the sarcomeres and the filaments is observed, and the number of the filaments in cross-sections also increases.

Considering filament lengths, the observations on *Drosophila* flight muscles suggest that the existing short filaments elongate through addition of myosin and actin monomers to their tips (Reedy and Beall 1993). Addition of actin monomers occurs mainly at the pointed ends of the thin filaments (Mardahl-Dumesnil and Fowler 2001). Increase in sarcomere width, in turn, could evidently occur only through addition of newly assembled filaments. Since no free filaments are observed around the myofibrils, the 'new' filaments must be assembling in close proximity to their final locations. Interestingly, the number of sarcomeres during the process remains unchanged, and their elongation is uniform along the myofibril. The mechanism stopping elongation of the filaments and the sarcomeres is not known. It appears that this event coincides with the ends of the elongating myofibrils nearing the cuticle surface (Reedy and Beall 1993). Overall, the observations seem to suggest that uniform synthesis of proteins along the nascent myofibrils is sufficient to ensure synchronous growth and length uniformity of the filaments and sarcomeres. Importantly, this pathway of myofibrillogenesis was observed only in the indirect flight muscles, and the other muscle types of *Drosophila* displayed different routes of assembly (Reedy and Beall 1993), suggesting the possibility of variations in the length controlling mechanisms.

One of the factors and proteins shown to affect filament length control in *Drosophila* flight muscles is *flightin*, a small (~30 kDa) accessory protein located on the surface of the thick filament backbone (Vigoreaux et al. 1993). In pupa of the flightin-null mutants of *Drosophila*, the thin and thick filaments, as well as the sarcomeres, are ~30 % longer than normal. However, after the beginning of contractile activity, thick filaments shorten and the sarcomere structure begins to progressively disintegrate (Reedy et al. 2000). Thus flightin appears to be important for both length regulation and structural stability of the filaments and sarcomeres. The role of flightin in the stability of the thick filaments was also illustrated by the significant increase in myosin monomer exchange observed in the *Drosophila* flightin-null mutants (Orfanos and Sparrow 2012).

Another protein of *Drosophila* flight muscles implicated, although indirectly, in the filament length regulation mechanism is obscurin (350–450 kDa), a member of the muscle Ig-family of proteins located at the M-lines (Burkart et al. 2007). Obscurin has an important role in defining symmetry of the thick filaments, i.e. in ensuring the central position of the bare zone (Katzemich et al. 2012). The loss of symmetry in the thick filaments, caused by reduction of obscurin expression, leads to a loss of length uniformity in the thin filaments.

Analysis of the actin- and myosin-null *Drosophila* mutants also showed that, in the absence of actin, only loose thick filament arrays could be assembled. In the absence of myosin, in turn, myofibril-like arrangements with aperiodically distributed Z-discs could be formed by continuous arrays of actin filaments (Beall et al. 1989). The loss in the latter case of regularity in the Z-disc positions reflects loss of length uniformity in the respective arrays of thin filaments caused by the absence of thick filaments, illustrating, as above, the inter-dependence of the myofilaments.

In the experiments specifically addressing the importance of contractility for sarcomerogenesis, transgenic myosin heavy chains lacking their actin-interacting

'head', the motor domain, were expressed in myosin-null indirect flight muscles of *Drosophila* (Cripps et al. 1999). It was found that the 'headless' myosin was capable of assembly into thick filaments, which were also able to integrate with actin filaments into sarcomere-like arrangements. However, in comparison with the control muscles, the lengths of the 'sarcomeres' as well as the thick filaments in the mutant muscles were highly variable (Cripps et al. 1999).

As for the roles of titin-like and nebulin-like proteins, flight muscles in the nebulin-repeat protein *Lasp*-lacking Drosophila mutants are shown to assemble seemingly normal periodic myofibrils (Fernandes and Schöck 2014). However, the sarcomeres, as well as the thin and thick filaments in the mutants, are measured to be ~12 % shorter than in the control muscles. Since *Lasp* is normally located in the Z-discs, the results suggest that Z-disc integrity and the correct interactions of the myofilaments with the Z-disc components comprise an important part in the mechanism of length regulation. The tight relationship between Z-discs (and M-lines) and the filament lengths is also illustrated by experiments with the inhibited expression of the *sallimus (Drosophila* titin) gene (Orfanos et al. 2015). Decreased amounts of the *sallimus* protein products result in characteristic changes in myofibril structure, namely, in restriction of the areas of normal Z- and M-line protein networks by the myofibril core, although overall diameter of myofibrils is the same as in wild-type fly. Within the core, the lengths of the sarcomeres and filaments appear normal. Beyond the core, at the myofibril periphery, the absence of the Z- and M-lines correlates with changes (increases) in length and the loss of symmetry of the thick filaments.

In summary, the indirect flight muscles of *Drosophila* present an example of myofilament length-regulation in the absence of a physical protein-ruler. The mechanism specifying the length and uniformity of the filaments appears to reflect a combined effect of different but inter-related contributors, as could be judged by the variety of factors that reveal their length-modulating effects. However, it is important to note that one of the factors, the acto-myosin interaction, is also the factor implicated in filament length control in vertebrate striated muscles.

10.7 Size Regulation in Non-muscle Biological Structures

The question of length regulation of muscle filaments is a part of a general and widely discussed problem of size control in biological systems (reviewed by Goehring and Hyman 2012; Levy and Heald 2012; Marshall 2015). The 'molecular ruler' hypothesis, or its variants, the 'template' and the 'tape measure', is one of the most popular ideas concerning linear structures. This mechanism was first proposed for assembly of the TMV virus, with the nucleic acid chain acting as a tape measure for the virus length (Caspar and Klug 1962). Since then, evidence has accumulated suggesting that similar length-defining mechanisms could also function in other linear viruses and various extended organelles of prokaryotes such as bacteriophage tails (Katsura 1987; Abuladze et al. 1994), the needles of bacterial injectisomes

(Journet et al. 2003; Wee and Hughes 2015), and the hook of bacterial filaments (Hughes 2012). In all these cases a protein/polypeptide, whose length is comparable with the final size of the polymer, is found in close contact with the polymerizing structure. The mechanism proposed to function in the bacteriophage tail suggests that, while the polymer tube grows around the ruler, the free end of the ruler inhibits binding of the 'terminator'-protein. As soon as the growing end of the polymer reaches the free end of the ruler, the protein-terminator binds to it and blocks further growth (Katsura 1990). In a recent report, a mechanism was also suggested explaining how, prior to its interactions with the polymerizing protein, the free part of the ruler could be protected from non-functional interactions by molecular chaperones (Xu et al. 2014), thus completing the mechanism. In the more complex systems of bacteria, however, some other factors, except molecular rulers, are also found to have length controlling effects. Some studies relate length control to accumulation of a defined amount of the monomer prior to polymerization (Makishima et al. 2001), others find a link with the rate and time of polymerization (Moriya et al. 2006), or with the stoichiometry of the proteins composing the structure (Marlovits et al. 2006). These studies illustrate an increase in complexity of the molecular ruler-related length controlling mechanisms in comparison with the simpler case of linear viruses (Aizawa 2012; Hughes 2012).

In the case of eukaryotes, only single examples of molecular rulers have been described, apart from muscle titin and nebulin. In particular, two protein complexes were proposed to determine the two characteristic periodicities, of ~96 nm (Oda et al. 2014) and ~24 nm (Owa et al. 2014), respectively, in *Chlamydomona*s flagella microtubular axonemes. The sizes of the rulers were shown to exactly match the observed periodicities. The ruler function of one of the proteins, however, has already been questioned, suggesting instead a structure stabilizing function (Oda et al. 2016). In the case of the flagella length, no protein with molecular ruler properties is known to be present there (reviewed by Marshall 2002, 2015; Wilson et al. 2008; Ishikawa and Marshall 2011). It must also be noted that the axoneme structure is more complex than, for example, the thin and thick filaments, although it is as regular as any linear homo-polymer. Also, unlike the static structures of prokaryotic linear organelles, flagellar axonemes are dynamic formations with continuing assembly-disassembly of the subunits at the tips (Marshall and Rosenbaum 2001; Song and Dentler 2001). Several factors have been shown to play a role in the length regulation mechanism, including intracellular levels of Ca^{2+} (Tuxhorn et al. 1998), the signaling proteins (Berman et al. 2003; Wilson and Lefebvre 2004: Tam et al. 2007) and the intra-flagellar (ITF) transport machinery. It is thought that ITF, which supplies the growing end with building blocks and removes the unbound material, establishes, at some point, a dynamic balance between the assembly-disassembly rates that stabilize the length (Marshall and Rosenbaum, 2001; Marshall et al. 2005; Engel et al. 2009). The IFT itself could be under control of Ca^{2+} and other signaling systems in the cell (Collingridge et al. 2013). Similar mechanisms based on subunit turnover and a dynamic balance of the assembly-disassembly rates, are also thought to function in specialized cellular protrusions, such as microtubular cilia (Ishikawa and Marshall 2011), as well as stereocilia, microvilli and some other analogous

formations, whose cores contain cross-linked actin bundles (Tilney et al. 1992; Manor and Kachar 2008). These mechanisms provide not only control of growth of the organelles, and thus of their core filaments, to a defined length, but also stabilize the organelles/filaments at this length. However, it must be noted that, in contrast to the molecular ruler-based mechanisms, the length in these dynamic structures is not exact, but fluctuates around a particular value.

10.8 Conclusion

Since titin and nebulin were proposed as 'molecular rulers' defining the lengths of the thick and thin filaments, respectively, it was generally assumed that they function as templates for the assembly of the filaments. Current research suggests that the proteins are unlikely to function in this way. It is also apparent that control of length, at least in eukaryotes, could be achieved without a physical ruler. Moreover, even where a protein of the size and structure compatible with the ruler function is present in the system, as in prokaryotes, its 'length-ruling' mechanism does not necessarily involves providing a pre-formed framework for the assembly of the polymer.

According to the current view, titin and nebulin interact with and stabilize actin and myosin polymers within the sarcomere structure. These interactions between the single-polypeptide proteins, each of an exact length, and the polymers, which could vary in size, result in strict uniformity of the mechanically stable segments of the filaments in the respective filament arrays in the sarcomeres. In this sense, titin and nebulin are rulers defining the lengths of mechanically stable segments of the filaments. The fact that the growing thin and thick filaments could reach the length of about correct size without the rulers, or be longer than the rulers, does not contradict the ruler hypotheses. These facts reflect the function and results of dynamic polymer assembly-disassembly mechanisms producing dynamic and fragile filaments, which cannot withstand tensions related to the contractile function of the sarcomeres, unless they are stabilized. From the point of view of a structure stabilizing function, the fact that the invertebrate striated muscles still contain single-polypeptide multidomain members of the titin and nebulin protein families, although of a smaller size, is likely to reflect a similar requirement for stabilization but of smaller segments of the filaments. The size and location of these proteins in the sarcomeres would then indicate the particular segments in the filaments where stability is specifically important for the muscles.

The suggested ruler-stabilizer function of titin and nebulin would not require the proteins to be arranged into the striated muscle sarcomere pattern prior to the arrangement of the filaments, as would be required were they protein-templates. This function is in agreement with the majority, if not all, studies related to the subject. However, it must be stressed that such a model does not simplify an answer to the question, how the interactions between titin/nebulin and the thick/thin filaments occur and are controlled during myofibrillogenesis. There are also significant differences in the way

nebulin and titin are bound to the respective filaments *in situ*. In the nebulin case, the entire molecule binds along the thin filament, and there are only two molecules bound symmetrically per thin filament. In the titin case, only a part of the molecule, the A-band part, binds to the thick filament, and it binds to only one half of the thick filament. There are also six molecules of titin, probably in three side-by-side dimers, that bind to each half of the filament (Al-Khayat et al. 2013). This difference in organization of the nebulin/thin filament and the titin/thick filament complexes may require a more intricate mechanism for the integration of titin with the sarcomeric myosin during myofibrillar assembly, especially if the mechanism would have to integrate myosin filaments pre-formed away from their final sarcomere locations. At present, it remains unclear whether the filaments are assembled distant from or directly at the sarcomere locations. Logically, assembly of the thick filaments directly in the future sarcomeres, and their concurrent interaction with titins, would provide a simpler mechanism. And this would also be in agreement with the observations suggesting involvement of microtubule-directed transport processes in titin-myosin assembly (Pizon et al. 2002, 2005), and/or with a possibility, although controversial, of co-translational assembly, evidenced by the accumulations of ribosomes in close proximity of the thick filaments in developing myofibrils (Larson et al.1969; 1973; Myklebust et al. 1978; Russel et al. 1992; Russel and Dix 1992; Gauthier and Mason-Savas 1993; Komiyama et al. 1993; Larsen and Sætersdal 1998).

To conclude, while the currently available information does not agree with the function of titin and nebulin as simple templates for the assembly of thick and thin filaments, it does not generally compromise the molecular ruler proposals. It rather indicates that the length-ruling function of the proteins is of a different type; their size defines the size of stabilized segments of the filaments.

References

Abuladze NK, Gingery M, Tsai J, Eiserling FA (1994) Tail length determination in bacteriophage T4. Virology 199:301–310

Aizawa S-I (2012) Mystery of FliK in length control of the flagellar hook. J Bacteriol 194:4798–4800

Alahiotis SN (1983) Heat shock proteins: a new view on the temperature compensation. Comp Biochem Physiol 75B:379–387

Albanèse V, Yam AY-W, Baughman J, Parnot C, Frydman J (2006) Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. Cell 124:75–88

Al-Khayat HA, Kensler RW, Squire JM, Marston SB, Morris EP (2013) Atomic model of the human cardiac muscle myosin filament. Proc Natl Acad Sci U S A 110:318–323

Allen ER, Pepe FA (1965) Ultrastructure of developing muscle cells in the chick embryo. Am J Anat 116:115–148

Altschuler GM, Klug DR, Willison KR (2005) Unfolding energetics of G-alpha-actin: a discrete intermediate can be re-folded to the native state by CCT. J Mol Biol 353:385–396

Anderson BR, Granzier HL (2012) Titin-based tension in the cardiac sarcomere: molecular origin and physiological adaptations. Prog Biophys Mol Biol 110(2–3,SI):204–217

Arimura T, Nakamura T, Hiroi S, Satoh M, Takahashi M, Ohbuchi N, Ueda K, Nouchi T, Yamaguchi N, Akai J, Matsumori A, Sasayama S, Kimura A (2000) Characterization of the human nebu-

lette gene: a polymorphism in an actin-binding motif is associated with nonfamilial idiopathic dilated cardiomyopathy. Hum Genet 107:440–451

Bang M-L, Chen J (2015) Roles of nebulin family members in the heart. Circ J 79:2081–2087

Bang M-L, Centner T, Fornoff F, Geach AJ, Gotthardt M, McNabb M, Witt CC, Labeit D, Gregorio CC, Granzier H, Labeit S (2001) The complete gene sequence of titin, Expression of an unusual ~700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. Circ Res 89:1065–1072

Bang M-L, Li X, Littlefield R, Bremner S, Thor A, Knowlton KU, Lieber RL, Chen J (2006) Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle. J Cell Biol 173:905–916

Barral JM, Hutagalung AH, Brinker A, Hartl FU, Epstein HF (2002) Role of the myosin assembly protein UNC-45 as a molecular chaperone for myosin. Science 295:669–671

Beall CJ, Sepanski MA, Fyrberg EA (1989) Genetic dissection of Drosophila myofibril formation: effects of actin and myosin heavy chain null alleles. Genes Dev 3:131–140

Begum S, Komiyama M, Toyota N, Obinata T, Maruyama K, Shimada Y (1998) Differentiation of muscle-specific proteins in chicken somites as studied by immunofluorescence microscopy. Cell Tissue Res 293:305–311

Berman SA, Wilson NF, Haas NA, Lefebvre PA (2003) A novel MAP kinase regulates flagellar length in *Chlamydomonas*. Curr Biol 13:1145–1149

Brennan C, Mangoli M, Dyer CEF, Ashworth R (2005) Acetylcholine and calcium signalling regulates muscle fibre formation in the *Zebrafish* embryo. J Cell Sci 118:5181–5190

Buck D, Hudson BD, Ottenheijm CAC, Labeit S, Granzier H (2010) Differential splicing of the large sarcomeric protein nebulin during skeletal muscle development. J Struct Biol 170:325–333

Bullard B, Linke WA, Leonard K (2002) Varieties of elastic protein in invertebrate muscles. J Muscle Res Cell Motil 23:435–447

Bullard B, Ferguson C, Minajeva A, Leake MC, Gautel M, Labeit D, Ding L, Labeit S, Horwitz J, Leonard KR, Linke WA (2004) Association of the chaperone alphaB-crystallin with titin in heart muscle. J Biol Chem 279:7917–7924

Bullard B, Burkart C, Labeit S, Leonard K (2005) The function of elastic proteins in the oscillatory contraction of insect flight muscle. J Muscle Res Cell Motil 26:479–485

Burgoyne T, Muhamad F, Luther PK (2008) Visualization of cardiac muscle thin filaments and measurement of their lengths by electron tomography. Cardiovasc Res 77:707–712

Burkart C, Qiu F, Brendel S, Benes V, Håååg P, Labeit S, Leonard K, Bullard B (2007) Modular proteins from the Drosophila sallimus (sls) gene and their expression in muscles with different extensibility. J Mol Biol 367:953–969

Cantino ME, Chew MWK, Luther PK, Morris E, Squire JM (2002) Structure and nucleotide-dependent changes of thick filaments in relaxed and rigor plaice fin muscle. J Struct Biol 137:164–175

Carroll SL, Horowits R (2000) Myofibrillogenesis and formation of cell contacts mediate the localization of N-RAP in cultured chick cardiomyocytes. Cell Motil Cytoskeleton 47:63–76

Carroll S, Lu S, Herrera AH, Horowits R (2004) N-RAP scaffolds I-Z-I assembly during myofibrillogenesis in cultured chick cardiomyocytes. J Cell Sci 117:105–114

Caspar DLD, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24

Castillo A, Nowak R, Littlefield KP, Fowler VM, Littlefield RS (2009) A nebulin ruler does not dictate thin filament lengths. Biophys J 96:1856–1865

Cazorla O, Freiburg A, Helmes M, Centner T, McNabb M, Wu Y, Trombitás K, Labeit S, Granzier H (2000) Differential expression of cardiac titin isoforms and modulation of cellular stiffness. Circ Res 86:59–67

Chen MJG, Wang K (1994) Conformational studies of a two-module fragment of nebulin and implications for actin association. Arch Biochem Biophys 310:310–317

Chen MJG, Shih CL, Wang K (1993) Nebulin as an actin zipper: a two-module nebulin fragment promotes actin nucleation and stabilizes actin filaments. J Biol Chem 268:20327–20334

Chitose R, Watanabe A, Asano M, Hanashima A, Sasano K, Bao Y, Maruyama K, Kimura S (2010) Isolation of nebulin from rabbit skeletal muscle and its interaction with actin. J Biomed Biotech 2010:108495. doi:10.1155/2010/108495

Chu M, Gregorio CC, Pappas CT (2016) Nebulin, a multi-functional giant. J Exp Biol 219:146–152

Colley NJ, Tokuyasu KT, Singer SJ (1990) The early expression of myofibrillar proteins in round postmitotic myoblasts of embryonic skeletal muscle. J Cell Sci 95:11–22

Collingridge P, Brownlee C, Wheeler GL (2013) Compartmentalized calcium signaling in cilia regulates intraflagellar transport. Curr Biol 23:2311–2318

Craig R, Offer G (1976) Axial arrangement of crossbridges in thick filaments of vertebrate skeletal muscle. J Mol Biol 102:325–332

Craig EA, Gambill BD, Nelson RJ (1993) Heat shock proteins: molecular chaperones of protein biogenesis. Microbiol Rev 57:402–414

Crawford GL, Horowits R (2011) Scaffolds and chaperones in myofibril assembly: putting the striations in striated muscle. Biophys Rev 3:25–32

Cripps RM, Suggs JA, Bernstein SI (1999) Assembly of thick filaments and myofibrils occurs in the absence of the myosin head. EMBO J 18:1793–1804

Davis JS (1981) Pressure-jump studies on the length-regulation kinetics of the self-assembly of myosin from vertebrate skeletal muscle into thick filament. Biochem J 197:309–314

Davis JS (1985) Kinetics and thermodynamics of the assembly of the parallel- and antiparallel-packed sections of synthetic thick filaments of skeletal myosin: pressure-jump study. Biochemist 24:5263–5269

Davis JS (1988) Assembly processes in vertebrate skeletal thick filament formation. Annu Rev Biophys Biophys Chem 17:217–239

De Deyne P (2000) Formation of sarcomeres in developing myotubes: role of mechanical stretch and contractile activation. Am J Physiol 279:C1801–C1811

Dessouky DA, Hibbs RG (1965) An electron microscope study of the development of the somatic muscle of the chick embryo. Am J Anat 116:523–566

Dlugosz AA, Antin PB, Nachmias VT, Holtzer H (1984) The relationship between stress fiber-like structures and nascent myofibrils in cultured cardiac myocytes. J Cell Biol 99:2268–2278

Donlin LT, Andresen C, Just S, Rudensky E, Pappas CT, Kruger M, Jacobs EY, Unger A, Zieseniss A, Dobenecker MW, Voelkel T, Chait BT, Gregorio CC, Rottbauer W, Tarakhovsky A, Linke WA (2012) Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. Genes Dev 26:114–119

Donner K, Sandbacka M, Lehtokari VL, Wallgren-Pettersson C, Pelin K (2004) Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. Eur J Hum Genet 12:744–751

Donner K, Nowak KJ, Aro M, Pelin K, Wallgren-Pettersson C (2006) Developmental and muscle-type-specific expression of mouse nebulin exons 127 and 128. Genomics 88:489–495

Dos Remedios CG, Chhabra D, Kekic M, Dedova IV, Tsubakihara M, Berry DA, Nosworthy NJ (2003) Actin binding proteins: regulation of cytoskeletal microfilaments. Physiol Rev 83:433–473

Du A, Sanger JM, Sanger JW (2008a) Cardiac myofibrillogenesis inside intact embryonic hearts. Dev Biol 318:236–246

Du SJ, Li H, Bian Y, Zhong Y (2008b) Heat-shock protein 90α1 is required for organized myofibril assembly in skeletal muscles of zebrafish embryos. Proc Natl Acad Sci U S A 105:554–559

Du SJ, Tan X, Zhang J (2014) SMYD Proteins: key regulators in skeletal and cardiac muscle development and function. Anat Rec 297:1650–1662

Dunn AY, Melville MW, Frydman J (2001) Review: cellular substrates of the eukaryotic chaperonin TRiC/CCT. J Struct Biol 135:176–184

Eddinger TJ (1998) Myosin heavy chain isoforms and dynamic contractile properties: skeletal versus smooth muscle. Comp Biochem Physiol B 119:425–434

Eggers DK, Welch WJ, Hansen WJ (1997) Complexes between nascent polypeptides and their molecular chaperones in the cytosol of mammalian cells. Mol Biol Cell 8:1559–1573

Ehler E, Rothen BM, Hämmerle SP, Komiyama M, Perriard J-C (1999) Myofibrillogenesis in the developing chicken heart: assembly of Z-disk, M-line and the thick filaments. J Cell Sci 112:1529–1539

Engel BD, Ludington WB, Marshall WF (2009) Intraflagellar transport particle size scales inversely with flagellar length: revisiting the balance-point length control model. J Cell Biol 187:81–89

Esham M, Bryan K, Milnes J, Holmes WB, Moncman CL (2007) Expression of nebulette during early cardiac development. Cell Motil Cytoskeleton 64:258–273

Etard C, Behra M, Fischer N, Hutcheson D, Geisler R, Strähle U (2007) The UCS factor Steif/Unc-45b interacts with the heat shock protein Hsp90a during myofibrillogenesis. Dev Biol 308:133–143

Etard C, Roostalu U, Strähle U (2008) Shuttling of the chaperones Unc45b and Hsp90a between the A band and the Z line of the myofibril. J Cell Biol 180:1163–1175

Fernandes I, Schöck F (2014) The nebulin repeat protein Lasp regulates I-band architecture and filament spacing in myofibrils. J Cell Biol 206:559–572

Ferrari MB, Spitzer NC (1999) Calcium signaling in the developing Xenopus myotome. Dev Biol 213:269–282

Ferrari MB, Podugu S, Eskew JD (2006) Assembling the myofibril – coordinating contractile cable construction with calcium. Cell Biochem Biophys 45:317–336

Friedrich BM, Buxboim A, Discher DE, Safran SA (2011) Striated acto-myosin fibers can reorganize and register in response to elastic interactions with the matrix. Biophys J 100:2706–2715

Fischman DA (1967) An electron microscope study of myofibril formation in embryonic chick skeletal muscle. J Cell Biol 32:557–575

Fowler VM, McKeown CR, Fischer RS (2006) Nebulin: does it measure up as a ruler? Curr Biol 16:R18–R20

Frydman J (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem 70:603–647

Fürst DO, Gautel M (1995) The anatomy of a molecular giant: how the sarcomere cytoskeleton is assembled from immunoglobulin superfamily molecules. J Mol Cell Cardiol 27:951–959

Fürst DO, Osborn M, Nave R, Weber K (1988) The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy: a map of ten nonrepetitive epitopes starting at the Z line extends close to the M line. J Cell Biol 106:1563–1572

Fürst DO, Nave R, Osborn M, Weber K (1989a) Repetitive titin epitopes with a 42 nm spacing coincide in relative position with known A band striations also identified by major myosin-associated proteins. An immunoelectron-microscopical study on myofibrils. J Cell Sci 94:119–125

Fürst DO, Osborn M, Weber K (1989b) Myogenesis in the mouse embryo: differential onset of expression of myogenic proteins and the involvement of titin in myofibril assembly. J Cell Biol 109:517–527

Fürst DO, Vinkemeier U, Weber K (1992) Mammalian skeletal muscle C-protein: purification from bovine muscle, binding to titin and the characterization of a full-length human cDNA. J Cell Sci 102:769–778

Gauthier GF, Mason-Savas A (1993) Ribosomes in the skeletal muscle filament lattice. Anat Rec 237:149–156

Gazda L, Pokrzywa W, Hellerschmied D, Löwe T, Forné I, Mueller-Planitz F, Hoppe T, Clausen T (2013) The myosin chaperone UNC-45 is organized in tandem modules to support myofilament formation in *C. elegans*. Cell 152:183–195

Geach TJ, Zimmerman LB (2010) Paralysis and delayed Z-disc formation in the *Xenopus tropicalis unc45b* mutant *dicky ticker*. BMC Dev Biol 10:N75

Geach TJ, Hirst EMA, Zimmerman LB (2015) Contractile activity is required for Z-disc sarcomere maturation in vivo. Genesis 53:299–307

Goehring NW, Hyman AA (2012) Organelle growth control through limiting pools of cytoplasmic components. Curr Biol 22:R330–R339

Gokhin DS, Lewis RA, McKeown CR, Nowak RB, Kim NE, Littlefield RS, Lieber RL, Fowler VM (2010) Tropomodulin isoforms regulate thin filament pointed-end capping and skeletal muscle physiology. J Cell Biol 189:95–109

Gokhin DS, Ochala J, Domenighetti AA, Fowler VM (2015) Tropomodulin 1 directly controls thin filament length in both wild-type and tropomodulin 4-deficient skeletal muscle. Development 142:4351–4362

Golenhofen N, Arbeiter A, Koob R, Drenckhahn D (2002) Ischemia-induced association of the stress protein αB-crystallin with I-band portion of cardiac titin. J Mol Cell Cardiol 34:309–319

Gotthardt M, Hammer RE, Hübner N, Monti J, Witt CC, McNabb M, Richardson JA, Granzier H, Labeit S, Herz J (2003) Conditional expression of mutant M-line titins results in cardiomyopathy with altered sarcomere structure. J Biol Chem 278:6059–6065

Gramlich M, Michely B, Krohne C, Heuser A, Erdmann B, Klaassen S, Hudson B, Magarin M, Kirchner F, Todiras M, Granzier H, Labeit S, Thierfelder L, Gerull B (2009) Stress-induced dilated cardiomyopathy in a knock-in mouse model mimicking human titin-based disease. J Mol Cell Cardiol 47:352–358

Greaser ML, Pleitner JM (2014) Titin isoform size is not correlated with thin filament length in rat skeletal muscle. Front Physiol 5:1–9

Gregorio CC, Weber A, Bondad M, Pennise CR, Fowler VM (1995) Requirement of pointed-end capping by tropomodulin to maintain actin filament length in embryonic chick cardiac myocytes. Nature 377:83–86

Gu X, Spitzer NC (1995) Distinct aspects of neuronal differentiation encoded by frequency of spontaneous Ca2+ transients. Nature 375:784–787

Guo W, Bharmal SJ, Esbona K, Greaser ML (2010) Titin diversity - alternative splicing gone wild. J Biomed Biotech 2010:753675. doi:10.1155/2010/753675

Hagopian M, Spiro D (1968) The filament lattice of cockroach thoracic muscle. J Cell Biol 36:433–442

Hanashima A, Kubokawa K, Kimura S (2009) Structure of the *amphioxus* nebulin gene and evolution of the nebulin family genes. Gene 443:76–82

Handel SE, Greaser ML, Schultz E, Wang SM, Bulinski JC, Lin JJC, Lessard JL (1991) Chicken cardiac myofibrillogenesis studied with antibodies specific for titin and the muscle and non-muscle isoforms of actin and tropomyosin. Cell Tissue Res 263:419–430

Harrington WF, Rodgers ME (1984) Myosin. Annu Rev Biochem 53:35–73

Harris BN, Li H, Terry M, Ferrari MB (2005) Calcium transients regulate titin organization during myofibrillogenesis. Cell Motil Cytoskeleton 60:129–139

Harrison RG, Lowey S, Cohen C (1971) Assembly of myosin. J Mol Biol 59:531–535

Hay ED (1963) The fine structure of differentiating muscle in the salamander tail. Zeitschrift fiir Zellforschung 59:6–34

Hendrick JP, Hartl FU (1993) Molecular chaperone functions of heat-shock proteins. Annu Rev Biochem 62:349–384

Higuchi H, Ishiwata S (1985) Disassembly kinetics of thick filaments in rabbit skeletal muscle fibers: Effects of ionic strength, Ca2+ concentration, pH, temperature, and cross-bridges on the stability of thick filament structure. Biophys J 47:267–275

Higuchi H, Funatsu T, Ishijima A, Okamura N, Ishiwata S (1986) Accumulated strain mechanism for length determination of thick filaments in skeletal muscle. I. Experimental bases. J Muscle Res Cell Motil 7:491–500

Hill CS, Duran S, Lin Z, Weber K, Holtzer H (1986) Titin and myosin, but not desmin, are linked during myofibrillogenesis in postmitotic mononucleated myoblasts. J Cell Biol 103:2185–2196

Holtzer H, Hijikata T, Lin ZX, Zhang ZQ, Holtzer S, Protasi F, Franzini-Armstrong C, Sweeney HL (1997) Independent assembly of 1.6 µm long bipolar MHC filaments and I-Z-I bodies. Cell Struct Funct 22:83–93

Hughes KT (2012) Flagellar hook length is controlled by a secreted molecular ruler. J Bacteriol 194:4793–4796

Huxley HE (1957) The double array of filaments in cross-striated muscle. J Biophys Biochem Cytol 3:631–648

Huxley HE (1963) Electron microscope studies on the structure of natural and synthetic protein filaments from striated muscle. J Mol Biol 7:281–308

Huxley HE (1967) Recent X-ray and electron microscope studies of striated muscle. J Gen Physiol 50:71–83

Huxley HE, Brown W (1967) The low-angle X-ray diagram of vertebrate striated muscle and its behaviour during contraction and rigor. J Mol Biol 30:383–434

Isaacs WB, Kim IS, Struve A, Fulton AB (1992) Association of titin and myosin heavy chain in developing skeletal muscle. Proc Natl Acad Sci U S A 89:7496–7500

Ishikawa H, Marshall WF (2011) Ciliogenesis: building the cell's antenna. Nat Rev Mol Cell Biol 12:222–234

Ishiwata S (1981) Melting from both ends of an A-band in a myofibril: observation with a phase-contrast microscope. J Biochem 89:1647–1650

Ishiwata S, Funatsu T (1985) Does actin bind to the ends of thin filaments in skeletal muscle? J Cell Biol 100:282–291

Ishiwata S, Muramatsu K, Higuchi H (1985) Disassembly from both ends of thick filaments in rabbit skeletal muscle fibers. Biophys J 47:257–266

Jin J-P, Wang K (1991) Nebulin as a giant actin-binding template protein in skeletal muscle sarcomere: Interaction of actin and cloned human nebulin fragments. FEBS Lett 281:93–96

Josephs R, Harrington WF (1966) Studies on the formation and physical chemical properties of synthetic myosin filaments. Biochemistry 5:3474–3487

Journet L, Agrain C, Broz P, Cornelis GR (2003) The needle length of bacterial injectisomes is determined by a molecular ruler. Science 302:1757–1760

Just S, Meder B, Berger IM, Etard C, Trano N, Patzel E, Hassel D, Marquart S, Dahme T, Vogel B, Fishman MC, Katus HA, Strähle U, Rottbauer W (2011) The myosin-interacting protein SMYD1 is essential for sarcomere organization. J Cell Sci 124:3127–3136

Kagawa M, Sato N, Obinata T (2006) Effects of BTS (N-benzyl-p-toluene sulphonamide), an inhibitor for myosin-actin interaction, on myofibrillogenesis in skeletal muscle cells in culture. Zool Sci 23:969–975

Kaminer B, Bell AL (1966) Synthetic myosin filaments. Science 151:323–324

Katsura I (1987) Determination of bacteriophage-lambda tail length by a protein ruler. Nature 327:73–75

Katsura I (1990) Mechanism of length determination in bacteriophage lambda tails. Adv Biophys 26:1–18

Katsura I, Noda H (1971) Studies on the formation and physical chemical properties of synthetic myosin filaments. J Biochem 69:219–229

Katzemich A, Kreisköther N, Alexandrovich A, Elliott C, Schöck F, Leonard K, Sparrow J, Bullard B (2012) The function of the M-line protein obscurin in controlling the symmetry of the sarcomere in the flight muscle of *Drosophila*. J Cell Sci 125:3367–3379

Kawamura M, Maruyama K (1970) Electron microscopic particle length of F-actin polymerized *in vitro*. J Biochem 67:437–457

Kazmierski ST, Antin PB, Witt CC, Huebner N, McElhinny AS, Labeit S, Gregorio CC (2003) The complete mouse nebulin gene sequence and the identification of cardiac nebulin. J Mol Biol 328:835–846

Kelly DE (1969) Myofibrillogenesis and Z-band differentiation. Anat Rec 163:403–426

Komiyama M, Zhou Z-H, Maruyama K, Shimada Y (1992) Spatial relationship of nebulin relative to other myofibrillar proteins during myogenesis in embryonic chick skeletal muscle cells *in vitro*. J Muscle Res Cell Motil 13:48–54

Komiyama M, Kouchi K, Maruyama K, Shimada Y (1993) Dynamics of actin and assembly of connectin (titin) during myofibrillogenesis in embryonic chick cardiac muscle cells in vitro. Dev Dyn 196:291–299

Kontrogianni-Konstantopoulos A, Ackermann MA, Bowman AL, Yap SV, Bloch RJ (2009) Muscle giants: molecular scaffolds in sarcomerogenesis. Physiol Rev 89:1217–1267

Kruger M, Wright J, Wang K (1991) Nebulin as a length regulator of thin filaments of vertebrate skeletal muscles: correlation of thin filament length, nebulin size, and epitope profile. J Cell Biol 115:97–107

Kuhlman PA (2005) Dynamic changes in the length distribution of actin filaments during polymerization can be modulated by barbed end capping proteins. Cell Motil Cytoskeleton 61:1–8

Labeit S, Kolmerer B (1995) Titins: giant proteins in charge of muscle ultrastructure and elasticity. Science 270:293–296

Labeit S, Kolmerer B (1996) The complete primary structure of human nebulin and its correlation to muscle structure. J Mol Biol 248:308–315

Labeit S, Barlow DP, Gautel M, Gibson T, Holt J, Hsieh C-L, Francke U, Leonard K, Wardale J, Whiting A, Trinick J (1990) A regular pattern of two types of 100-residue motif in the sequence of titin. Nature 345:273–276

Labeit S, Gibson T, Lakey A, Leonard K, Zeviani M, Knight P, Wardale J, Trinick J (1991) Evidence that nebulin is a protein-ruler in muscle thin filaments. FEBS Lett 282:313–316

Labeit S, Gautel M, Lakey A, Trinick J (1992) Towards a molecular understanding of titin. EMBO J 11:1711–1716

Lahmers S, Wu Y, Call DR, Labeit S, Granzier H (2004) Developmental control of titin isoform expression and passive stiffness in fetal and neonatal myocardium. Circ Res 94:505–513

Larsen TH, Sætersdal T (1998) Translocation of 60S ribosomal subunit in spreading cardiac myocytes. J Histochem Cytochem 46:963–969

Larson PF, Hudgson P, Walton JN (1969) Morphological relationship of polyribosomes and myosin filaments in developing and regenerating skeletal muscle. Nature 222:1168–1169

Larson PF, Fulthorpe JJ, Hudgson P (1973) Alignment of polysomes along myosin filaments in developing myofibrils. J Anat 116:327–334

Lazarides E, Burridge K (1975) alpha-Actinin: immunofluorescent localization of a muscle structural protein in nonmuscle cells. Cell 6:289–298

Levy DL, Heald R (2012) Mechanisms of intracellular scaling. Annu Rev Cell Dev Biol 28:113–135

Li S, Guo W, Schmitt BM, Greaser ML (2012) Comprehensive analysis of titin protein isoform and alternative splicing in normal and mutant rats. J Cell Biol 113:1265–1273

Li H, Zhong Y, Wang Z, Gao J, Xu J, Chu W, Zhang J, Fang S, Du SJ (2013) Smyd1b is required for skeletal and cardiac muscle function in zebrafish. Mol Biol Cell 24:3511–3521

Lin ZX, Schultheiss T, Choi J, Holtzer S, Dilullo C, Fischman DA, Holtzer H (1994) Sequential appearance of muscle-specific proteins in myoblasts as a function of time after cell division – evidence for a conserved myoblast differentiation program in skeletal muscle. Cell Motil Cytoskeleton 29:1–19

Littlefield R, Fowler VM (2002) Measurement of thin filament lengths by distributed deconvolution analysis of fluorescence images. Biophys J 82:2548–2564

Liversage AD, Holmes D, Knight PJ, Tskhovrebova L, Trinick J (2001) Titin and the sarcomere symmetry paradox. J Mol Biol 305:401–409

Llorca O, McCormack EA, Hynes G, Grantham J, Cordell J, Carrascosa JL, Willison KR, Fernandez JJ, Valpuesta JM (1999) Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. Nature 402:693–696

Lorenzon P, Giovannelli A, Ragozzino D, Eusebi F, Ruzzier F (1997) Spontaneous and repetitive calcium transients in C2C12 mouse myotubes during in vitro myogenesis. Eur J Neurosci 9:800–808

Lu MH, DiLullo C, Schultheiss T, Holtzer S, Murray JM, Choi J, Fischman DA, Holtzer H (1992) The vinculin/sarcomeric-alpha-actinin/alpha-actin nexus in cultured cardiac myocytes. J Cell Biol 117:1007–1022

Lu S, Carroll SL, Herrera AH, Ozanne B, Horowits R (2003) New N-RAP-binding partners α-actinin, filamin and Krp1 detected by yeast two-hybrid screening: implications for myofibril assembly. J Cell Sci 116:2169–2178

Luo G, Herrera AH, Horowits R (1999) Molecular interactions of N-RAP, a nebulin-related protein of striated muscle myotendon junctions and intercalated disks. Biochemistry 38:6135–6143

Makishima S, Komoriya K, Yamaguchi S, Aizawa S-I (2001) Length of the flagellar hook and the capacity of the Type III export apparatus. Science 291:2411–2413

Manisastry SM, Zaal KJM, Horowits R (2009) Myofibril assembly visualized by imaging N-RAP, alpha-actinin, and actin in living cardiomyocytes. Exp Cell Res 315:2126–2139

Manor U, Kachar B (2008) Dynamic length regulation of sensory stereocilia. Semin Cell Dev Biol 19:502–510

Mardahl-Dumesnil M, Fowler VM (2001) Thin filaments elongate from their pointed ends during myofibril assembly in *Drosophila* indirect flight muscle. J Cell Biol 155:1043–1053

Marlovits TC, Kubori T, Lara-Tejero M, Thomas D, Unger VM, Galán JE (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. Nature 441:637–640

Marshall WF (2002) Size control in dynamic organelles. Trends Cell Biol 12:414–419

Marshall WF (2015) How cells measure length on subcellular scales. Trends Cell Biol 25:760–768

Marshall WF, Rosenbaum JL (2001) Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control. J Cell Biol 155:405–414

Marshall WF, Qin HM, Brenni MR, Rosenbaum JL (2005) Flagellar length control system: testing a simple model based on intraflagellar transport and turnover. Mol Biol Cell 16:270–278

Maruyama K, Natori R, Nonomura Y (1976) New elastic protein from muscle. Nature 262:58–60

Maruyama K, Matsuno A, Higuchi H, Shimaoka S, Kimura S, Shimizu T (1989) Behaviour of connectin (titin) and nebulin in skinned muscle fibres released after extreme stretch as revealed by immunoelectron microscopy. J Muscle Res Cell Motil 10:350–359

Mastrototaro G, Liang X, Li X, Carullo P, Piroddi N, Tesi C, Gu Y, Dalton ND, Peterson KL, Poggesi C, Sheikh F, Chen J, Bang M-L (2015) Nebulette knockout mice have normal cardiac function, but show Z-line widening and up-regulation of cardiac stress markers. Cardiovasc Res 107:216–225

McElhinny AS, Kolmerer B, Fowler VM, Labeit S, Gregorio CC (2001) The N-terminal end of nebulin interacts with tropomodulin at the pointed ends of the thin filaments. J Biol Chem 276:583–592

McElhinny AS, Schwach C, Valichnac M, Mount-Patrick S, Gregorio CC (2005) Nebulin regulates the assembly and lengths of the thin filaments in striated muscle. J Cell Biol 170:947–957

Miller G, Musa H, Gautel M, Peckham M (2003) A targeted deletion of the C-terminal end of titin, including the titin kinase domain, impairs myofibrillogenesis. J Cell Sci 116:4811–4819

Millevoi S, Trombitás K, Kolmerer B, Kostin S, Schaper J, Pelin K, Granzier H, Labeit S (1998) Characterization of nebulette and nebulin and emerging concepts of their roles for vertebrate Z-discs. J Mol Biol 282:111–123

Miyahara M, Noda H (1980) Interaction of C-protein with myosin. J Biochem 87:1413–1420

Moncman CL, Wang K (1995) Nebulette: a 107 kD nebulin-like protein in cardiac muscle. Cell Motil Cytoskeleton 32:205–225

Moncman CL, Wang K (1996) Assembly of nebulin into the sarcomeres of avian skeletal muscle. Cell Motil Cytoskeleton 34:167–184

Moos C, Offer G, Starr R, Bennett P (1975) Interaction of C-Protein with myosin, myosin rod and light meromyosin. J Mol Biol 97:1–9

Moriya N, Minamino T, Hughes KT, Macnab RM, Namba K (2006) The type III flagellar export specificity switch is dependent on FliK ruler and a molecular clock. J Mol Biol 359:466–477

Musa H, Meek S, Gautel M, Peddie D, Smith AJH, Peckham M (2006) Targeted homozygous deletion of M-band titin in cardiomyocytes prevents sarcomere formation. J Cell Sci 119:4322–4331

Myhre JL, Pilgrim DB (2012) At the start of the sarcomere: a previously unrecognized role for myosin chaperones and associated proteins during early myofibrillogenesis. Biochem Res Int 712315.doi:10.1155/2012/712315

Myhre JL, Hills JA, Prill K, Wohlgemuth SL, Pilgrim DB (2014) The titin A-band rod domain is dispensable for initial thick filament assembly in zebrafish. Dev Biol 387:93–108

Myklebust R, Sœtersdal TS, Engedal H, Ulstein M, Ødegården S (1978) Ultrastructural studies on the formation of myofilaments and myofibrils in the human embryonic and adult hypertrophied heart. Anat Embryol 152:127–140

Nagandla H, Lopez S, Yu W, Rasmussen TL, Tucker HO, Schwartz RJ, Stewart MD (2016) Defective myogenesis in the absence of the muscle-specific lysine methyltransferase SMYD1. Dev Biol 410:86–97

Nave R, Fürst DO, Weber K (1989) Visualization of the polarity of isolated titin molecules: a single globular head on a long thin rod as the M band anchoring domain? J Cell Biol 109:2177–2187

Nwe TM, Maruyama K, Shimada Y (1999) Relation of nebulin and connectin (titin) to dynamics of actin in nascent myofibrils of cultured skeletal muscle cells. Exp Cell Res 252:33–40

O'Brien EJ, Bennett PM, Hanson J (1971) Optical diffraction studies of myofibrillar structure. Philos Trans R Soc Lond B 261:201–208

Obinata T, Yamamoto M, Maruyama K (1966) The identification of randomly formed thin filaments in differentiating muscle cells of the chick embryo. Dev Biol 14:192–213

Oda T, Yanagisawa H, Kamiya R, Kikkawa M (2014) A molecular ruler determines the repeat length in eukaryotic cilia and flagella. Science 346:857–860

Oda T, Abe T, Yanagisawa H, Kikkawa M (2016) Docking-complex-independent alignment of *Chlamydomonas* outer dynein arms with 24-nm periodicity *in vitro*. J Cell Sci 129:1547–1551

Ogut O, Hossain MM, Jin JP (2003) Interactions between nebulin-like motifs and thin filament regulatory proteins. J Biol Chem 278:3089–3097

Opitz CA, Leake MC, Makarenko I, Benes V, Linke WA (2004) Developmentally regulated switching of titin size alters myofibrillar stiffness in the perinatal heart. Circ Res 94:967–975

Orfanos Z, Sparrow JC (2012) Myosin isoform switching during assembly of the *Drosophila* flight muscle thick filament lattice. J Cell Sci 126:139–148

Orfanos Z, Leonard K, Elliott C, Katzemich A, Bullard B, Sparrow J (2015) *Sallimus* and the dynamics of sarcomere assembly in *Drosophila* flight muscles. J Mol Biol 427:2151–2158

Ottenheijm CAC, Knottnerus AM, Buck D, Luo X, Greer K, Hoying A, Labeit S, Granzier H (2009) Tuning passive mechanics through differential splicing of titin during skeletal muscle development. Biophys J 97:2277–2286

Owa M, Furuta A, Usukura J, Arisaka F, King SM, Witman GB, Kamiya R, Wakabayashi K (2014) Cooperative binding of the outer arm-docking complex underlies the regular arrangement of outer arm dynein in the axoneme. Proc Natl Acad Sci U S A 111:9461–9466

Page SG, Huxley HE (1963) Filament lengths in striated muscle. J Cell Biol 19:369–390

Pappas CT, Krieg PA, Gregorio CC (2010) Nebulin regulates actin filament lengths by a stabilization mechanism. J Cell Biol 189:859–870

Pappas CT, Bliss KT, Zieseniss A, Gregorio CC (2011) The nebulin family: an actin support group. Trends Cell Biol 21:29–37

Parsell DA, Lindquist S (1993) The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu Rev Genet 27:437–496

Pelham H (1985) Activation of heat-shock genes in eukaryotes. Trends Genet 1:31–35

Peng J, Raddatz K, Labeit S, Granzier H, Gotthardt M (2005) Muscle atrophy in titin M-line deficient mice. J Muscle Res Cell Motil 26:381–388

Peng J, Raddatz K, Molkentin JD, Wu Y, Labeit S, Granzier H, Gotthardt M (2007) Cardiac hypertrophy and reduced contractility in hearts deficient in the titin kinase region. Circulation 115:743–751

Pepe FA (1967) The myosin filament. I Structural organization from antibody staining observed in electron microscopy. J Mol Biol 37:203–225

Person V, Kostin S, Suzuki K, Labeit S, Schaper J (2000) Antisense oligonucleotide experiments elucidate the essential role of titin in sarcomerogenesis in adult rat cardiomyocytes in long-term culture. J Cell Sci 113:3851–3859

Pfuhl M, Winder SJ, Pastore A (1994) Nebulin, a helical actin-binding protein. EMBO J 13:1782–1789

Pierobon-Bormioli S, Betto R, Salviati G (1989) The organization of titin (connectin) and nebulin in the sarcomeres: an immunocytolocalization study. J Muscle Res Cell Motil 10:446–456

Pinset-Härström I (1985) MgATP specifically controls in vitro self-assembly of vertebrate skeletal myosin in the physiological pH range. J Mol Biol 182:159–172

Pizon V, Iakovenko A, van der Ven PFM, Kelly R, Fatu C, Fürst DO, Karsenti E, Gautel M (2002) Transient association of titin and myosin with microtubules in nascent myofibrils directed by the MURF2 RING-finger protein. J Cell Sci 115:4469–4482

Pizon V, Gerbal F, Diaz CC, Karsenti E (2005) Microtubule-dependent transport and organization of sarcomeric myosin during skeletal muscle differentiation. EMBO J 24:3781–3792

Pokrzywa W, Hoppe T (2013) Chaperoning myosin assembly in muscle formation and aging. Worm 2(3):e25644. doi:10.4161/worm.25644

Ramachandran I, Terry M, Ferrari MB (2003) Skeletal muscle myosin cross-bridge cycling is necessary for myofibrillogenesis. Cell Motil Cytoskeleton 55:61–72

Reedy MC, Beall C (1993) Ultrastructure of developing flight muscle in *Drosophila*. I. Assembly of myofibrils. Dev Biol 160:443–465

Reedy MC, Bullard B, Vigoreaux JO (2000) Flightin is essential for thick filament assembly and sarcomere stability in *Drosophila* flight muscles. J Cell Biol 151:1483–1499

Reisler E, Smith C, Seegan G (1980) Myosin mini-filaments. J Mol Biol 143:129–145

Rhee D, Sanger JM, Sanger JW (1994) The premyofibril: evidence for its role in myofibrillogenesis. Cell Motil Cytoskeleton 28:1–24

Richter K, Haslbeck M, Buchner J (2010) The heat shock response: life on the verge of death. Mol Cell 40:253–266

Ringkob TP, Swartz DR, Greaser ML (2004) Light microscopy and image analysis of thin filament lengths utilizing dual probes on beef, chicken, and rabbit myofibrils. J Anim Sci 82:1445–1453

Robinson TF, Winegrad S (1977) Variation of thin filament length in heart muscle. Nature 267:74–75

Robinson TF, Winegrad S (1979) The measurement and dynamic implications of thin filament lengths in heart muscle. J Physiol 286:607–619

Rudy DE, Yatskievych TA, Antin PB, Gregorio CC (2001) Assembly of thick, thin, and titin filaments in chick precardiac explants. Dev Dyn 221:61–71

Russel B, Dix DJ (1992) Mechanisms for intracellular distribution of mRNA: in situ hybridization studies in muscle. Am J Physiol 262:C1–C8

Russel B, Wenderoth MP, Goldspink PH (1992) Remodeling of myofibrils: subcellular distribution of myosin heavy chain mRNA and protein. Am J Physiol 262:R339–R345

Samarel AM (2005) Costameres, focal adhesions, and cardiomyocyte mechanotransduction. Am J Physiol 289:H2291–H2301

Sanger JW, Kang S, Siebrands CC, Freeman N, Du A, Wang J, Stout AL, Sanger JM (2005) How to build a myofibril. J Muscle Res Cell Motil 26:343–354

Sanger JW, Wang J, Fan Y, White J, Sanger JM (2010) Assembly and dynamics of myofibrils. J Biomed Biotech 2010:858606. doi:10.1155/2010/858606

Schaart G, Viebahn C, Langmann W, Ramaekers F (1989) Desmin and titin expression in early postimplantation mouse embryos. Development 107:585–596

Schultheiss T, Lin Z, Lu MH, Murray J, Fischman DA, Weber K, Masaki T, Imamura M, Holtzer H (1990) Differential distribution of subsets of myofibrillar proteins in cardiac nonstriated and striated myofibrils. J Cell Biol 110:1159–1172

Seeley M, Huang W, Chen Z, Wolff WO, Lin X, Xu X (2007) Depletion of zebrafish titin reduces cardiac contractility by disrupting the assembly of Z-discs and A-bands. Circ Res 100:238–245

Shih CL, Chen MJG, Linse K, Wang K (1997) Molecular contacts between nebulin and actin: cross-linking of nebulin modules to the N-terminus of actin. Biochemist 36:1814–1825

Shimada Y, Fischman DA, Moscona AA (1967) The fine structure of embryonic chick skeletal muscle cells differentiated *in vitro*. J Cell Biol 35:445–453

Sjöström M, Squire JM (1977) Fine structure of the A-band in cryo-sections: the structure of the A-band of human skeletal muscle fibres from ultra-thin cryo-sections negatively stained. J Mol Biol 109:49–68

Smith DA, Carland CR, Guo Y, Bernstein SI (2014) Getting folded: chaperone proteins in muscle development, maintenance and disease. Anat Rec 297:1637–1649

Soeno Y, Shimada Y, Obinata T (1999) BDM (2,3-butanedione monoxime), an inhibitor of myosin-actin interaction, suppresses myofibrillogenesis in skeletal muscle cells in culture. Cell Tissue Res 295:307–316

Sœtersdal T, Engedal H, Lie R, Myklebust R (1980) On the origin of Z-band material and myofilaments in myoblasts from the human atrial wall. Cell Tissue Res 207:21–29

Song L, Dentler WL (2001) Flagellar protein dynamics in *Chlamydomonas*. J Biol Chem 276:29754–29763

Sonoda M, Kimura S, Moriya H, Shimada Y, Maruyama K (1990) Molecular shape of alpha-connectin, an elastic filamentous protein of skeletal muscle. Proc Jpn Acad Ser B 66:213–216

Sosnicki AA, Loesser KE, Rome LC (1991) Myofilament overlap in swimming carp. I. Myofilament lengths of red and white muscle. Am J Physiol 260:C283–C288

Soteriou A, Gamage M, Trinick J (1993) A survey of interactions made by the giant protein titin. J Cell Sci 104:119–123

Sparrow JC, Schöck F (2009) The initial steps of myofibril assembly: integrins pave the way. Nat Rev Mol Cell Biol 10:293–298

Spiess C, Meyer AS, Reissmann S, Frydman J (2004) Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. Trends Cell Biol 14:598–604

Spiro D, Sonnenblick EH (1964) Comparison of the ultrastructural basis of the contractile process in heart and skeletal muscle. Circ Res 15:14–37

Spitzer NC (1994) Spontaneous Ca^{2+} spikes and waves in embryonic neurons: signaling systems for differentiation. TINS 17:115–118

Spotnitz HM, Sonnenblick EH, Spiro D (1966) Relation of ultrastructure to function in the intact heart: sarcomere structure relative to pressure volume curves of intact left ventricles of dog and cat. Circ Res 18:49–66

Squire JM, Roessle M, Knupp C (2004) New X-ray diffraction observations on vertebrate muscle: organisation of C-protein (MyBP-C) and troponin and evidence for unknown structures in the vertebrate A-band. J Mol Biol 343:1345–1363

Srikakulam R, Liu L, Winkelmann DA (2008) Unc45b forms a cytosolic complex with Hsp90 and targets the unfolded myosin motor domain. PLoS One 3(5):e2137. doi:10.1371/journal.pone.0002137

Sternlicht H, Farr GW, Sternlicht ML, Driscoll JK, Willison K, Yaffe MB (1993) The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin *in vivo*. Proc Natl Acad Sci U S A 90:9422–9426

Tam LW, Wilson NF, Lefebvre PA (2007) A CDK-related kinase regulates the length and assembly of flagella in *Chlamydomonas*. J Cell Biol 176:819–829

Tan X, Rotllant J, Li H, DeDeyne P, Du SJ (2006) SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos. Proc Natl Acad Sci U S A 103:2713–2718

Thulasiraman V, Yang CF, Frydman J (1999) *In vivo* newly translated polypeptides are sequestered in a protected folding environment. EMBO J 18:85–95

Tilney LG, Tilney MS, DeRosier DJ (1992) Actin filaments, sterocilia, and hair cells: how cells count and measure. Annu Rev Cell Biol 8:257–274

Tokuyasu KT, Maher PA (1987a) Immunocytochemical studies of cardiac myofibrillogenesis in early chick embryos. I. Presence of immunofluorescent titin spots in premyofibril stages. J Cell Biol 105:2781–2793

Tokuyasu KT, Maher PA (1987b) Immunocytochemical studies of cardiac myofibrillogenesis in early chick embryos.II. Generation of alpha-actinin dots within titin spots at the time of the first myofibril formation. J Cell Biol 105:2795–2801

Tondeleir D, Vandamme D, Vandekerckhove J, Ampe C, Lambrechts A (2009) Actin isoform expression patterns during mammalian development and in pathology: insights from mouse models. Cell Motil Cytoskeleton 66:798–815

Tonino P, Pappas CT, Hudson BD, Labeit S, Gregorio CC, Granzier H (2010) Reduced myofibrillar connectivity and increased Z-disk width in nebulin-deficient skeletal muscle. J Cell Sci 123:384–391

Traeger L, Goldstein MA (1983) Thin filaments are not of uniform length in rat skeletal muscle. J Cell Biol 96:100–103

Trinick J (1992) Molecular rulers in muscle? Curr Biol 2:75–77

Trinick J, Cooper J (1980) Sequential disassembly of vertebrate muscle thick filaments. J Mol Biol 141:315–321

Trinick J, Knight P, Whiting A (1984) Purification and properties of native titin. J Mol Biol 180:331–356

Trombitás K, Redkar A, Centner T, Wu Y, Labeit S, Granzier H (2000) Extensibility of isoforms of cardiac titin: variation in contour length of molecular subsegments provides a basis for cellular passive stiffness diversity. Biophys J 79:3226–3234

Trombitás K, Wu Y, Labeit D, Granzier H (2001) Cardiac titin isoforms are coexpressed in the half-sarcomere and extend independently. Am J Physiol 281:H1793–H1799

Tskhovrebova L, Trinick J (2001) Flexibility and extensibility in the titin molecule: analysis of electron microscope data. J Mol Biol 310:755–771

Tskhovrebova L, Trinick J (2003) Titin: Properties and family relationship. Nat Rev Mol Cell Biol 4:679–689

Tskhovrebova L, Walker ML, Grossmann JG, Khan GN, Baron A, Trinick J (2010) Shape and flexibility in the titin 11-domain super-repeat. J Mol Biol 397:1092–1105

Tu MK, Levin JB, Hamilton AM, Borodinsky LN (2016) Calcium signaling in skeletal muscle development, maintenance and regeneration. Cell Calcium 59:91–97

Tuxhorn J, Daise T, Dentler WL (1998) Regulation of flagellar length in *Chlamydomonas*. Cell Motil Cytoskeleton 40:133–146

van der Loop FTL, Schaart G, Langmann W, Ramaekers FCS, Viebahn C (1992) Expression and organization of muscle specific proteins during the early developmental stages of the rabbit heart. Anat Embryol 185:439–450

van der Loop FTL, van Eys GJJ, Schaart G, Ramaekers FCS (1996a) Titin expression as an early indication of heart and skeletal muscle differentiation in vitro. Developmental re-organisation in relation to cytoskeletal constituents. J Muscle Res Cell Motil 17:23–36

van der Loop FTL, van der Ven PFM, Fürst DO, Gautel M, van Eys GJJ, Ramaekers FCS (1996b) Integration of titin into sarcomeres of cultured differentiating human skeletal muscle cells. Eur J Cell Biol 69:301–307

van der Ven PFM, Ehler E, Perriard J-C, Fürst DO (1999) Thick filament assembly occurs after the formation of a cytoskeletal scaffold. J Muscle Res Cell Motil 20:569–579

van der Ven PFM, Bartsch JW, Gautel M, Jockusch H, Fürst DO (2000) A functional knock-out of titin results in defective myofibril assembly. J Cell Sci 113:1405–1414

Vigoreaux JO, Saide JD, Valgeirsdottir K, Pardue ML (1993) Flightin, a novel myofibrillar protein of *Drosophila* stretch-activated muscles. J Cell Biol 121:587–598

Voelkel T, Andresen C, Unger A, Just S, Rottbauer W, Linke WA (2013) Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function. Biochim Biophys Acta/Mol Cell Res 1833:812–822

Walker SM, Schrodt GR (1974) I Segment lengths and thin filament periods in skeletal muscle fibers of the rhesus monkey and the human. Anat Rec 178:63–81

Wang K, Williamson CL (1980) Identification of an N2 line protein of striated muscle. Proc Natl Acad Sci U S A 77:3254–3258

Wang K, Wright J (1988) Architecture of the sarcomere matrix of skeletal muscle: immunoelectron microscopic evidence that suggests a set of parallel inextensible nebulin filaments anchored at the Z Line. J Cell Biol 107:2199–2212

Wang K, McClure J, Tu A (1979) Titin: major myofibrillar components of striated muscle. Proc Natl Acad Sci U S A 76:3698–3702

Wang K, Ramirez-Mitchell R, Palter D (1984) Titin is an extraordinarily long, flexible, and slender myofibrillar protein. Proc Natl Acad Sci U S A 81:3685–3689

Wang S-M, Greaser ML, Schultz E, Bulinski JC, Lin JJ-C, Lessard JL (1988) Studies on cardiac myofibrillogenesis with antibodies to titin, actin, tropomyosin, and myosin. J Cell Biol 107:1075–1083

Wang K, McCarter R, Wright J, Beverly J, Ramirez-Mitchell R (1991) Regulation of skeletal muscle stiffness and elasticity by titin isoforms: a test of the segmental extension model of resting tension. Proc Natl Acad Sci U S A 88:7101–7105

Wang K, Knipfer M, Huang Q-Q, van Heerden A, Hsu LC-L, Gutierrez G, Quian X-L, Stedman H (1996) Human skeletal muscle nebulin sequence encodes a blueprint for thin filament architecture. Sequence motifs and affinity profiles of tandem repeats and terminal SH3. J Biol Chem 271:4304–4314

Warren CM, Krzesinski PR, Campbell KS, Moss RL, Greaser ML (2004) Titin isoform changes in rat myocardium during development. Mech Dev 121:1301–1312

Weber A, Pennise CR, Babcock GG, Fowler VM (1994) Tropomodulin caps the pointed ends of actin filaments. J Cell Biol 127:1627–1635

Wee DH, Hughes KT (2015) Molecular ruler determines needle length for the Salmonella Spi-1 injectisome. Proc Natl Acad Sci U S A 112:4098–4103

Weinert S, Bergmann N, Luo X, Erdmann B, Gotthardt M (2006) M line–deficient titin causes cardiac lethality through impaired maturation of the sarcomere. J Cell Biol 173:559–570

Whiting A, Wardale J, Trinick J (1989) Does titin regulate the length of muscle thick filaments? J Mol Biol 205:263–268

Wilson NF, Lefebvre PA (2004) Regulation of flagellar assembly by glycogen synthase kinase 3 in *Chlamydomonas reinhardtii*. Eukaryot Cell 3:1307–1319

Wilson NF, Iyer JK, Buchheim JA, Meek W (2008) Regulation of flagellar length in *Chlamydomonas*. Semin Cell Dev Biol 19:494–501

Witt CC, Burkart C, Labeit D, McNabb M, Wu Y, Granzier H, Labeit S (2006) Nebulin regulates thin filament length, contractility, and Z-disk structure *in vivo*. EMBO J 25:3843–3855

Wohlgemuth SL, Crawford BD, Pilgrim DB (2007) The myosin co-chaperone UNC-45 is required for skeletal and cardiac muscle function in zebrafish. Dev Biol 303:483–492

Wright J, Huang Q-Q, Wang K (1993) Nebulin is a full-length template of actin filaments in the skeletal muscle sarcomere: an immunoelectron microscopic study of its orientation and span with site-specific monoclonal antibodies. J Muscle Res Cell Motil 14:476–483

Xu X, Meiler SE, Zhong TP, Mohideen M, Crossley DA, Burggren WW, Fischman MC (2002) Cardiomyopathy in zebrafish due to mutation in an alternatively spliced exon of *titin*. Nat Genet 30:205–209

Xu J, Hendrix RW, Duda RL (2014) Chaperone–protein interactions that mediate assembly of the bacteriophage lambda tail to the correct length. J Mol Biol 426:1004–1018

Yadavalli VK, Forbes JG, Wang K (2009) Nanomechanics of full-length nebulin: an elastic strain gauge in the skeletal muscle sarcomere. Langmuir 25:7496–7505

Yamashiro S, Gokhin DS, Kimura S, Nowak RB, Fowler VM (2012) Tropomodulins: pointed-end capping proteins that regulate actin filament architecture in diverse cell types. Cytoskeleton 69:337–370

Yang J, Shih YH, Xu X (2014) Understanding cardiac sarcomere assembly with zebrafish genetics. Anat Rec/Adv Integr Evol Biol 297:1681–1693

Yasuda K, Fujita H, Fujiki Y, Ishiwata S (1994) Length regulation of thin filaments without nebulin. Proc Japan Acad Ser B/Phys Biol Sci 70:151–156

Zhang JQ, Luo G, Herrera AH, Paterson B, Horowits R (1996) cDNA cloning of mouse nebulin: evidence that the nebulin-coding sequence is highly conserved among vertebrates. Eur J Biochem 239:835–841

Zoghbi ME, Woodhead JL, Moss RL, Craig R (2008) Three-dimensional structure of vertebrate cardiac muscle myosin filaments. Proc Natl Acad Sci U S A 105:2386–2390

Chapter 11
Myosin and Actin Filaments in Muscle: Structures and Interactions

John M. Squire, Danielle M. Paul, and Edward P. Morris

Contents

Abstract In the last decade, improvements in electron microscopy and image processing have permitted significantly higher resolutions to be achieved (sometimes <1 nm) when studying isolated actin and myosin filaments. In the case of actin filaments the changing structure when troponin binds calcium ions can be followed

J.M. Squire (✉) • D.M. Paul
Muscle Contraction Group, School of Physiology, Pharmacology and Neuroscience,
University of Bristol, Bristol BS8 1TD, UK
e-mail: j.squire@imperial.ac.uk

E.P. Morris
Division of Structural Biology, The Institute of Cancer Research, London SW3 6JB, UK

using electron microscopy and single particle analysis to reveal what happens on each of the seven non-equivalent pseudo-repeats of the tropomyosin α-helical coiled-coil. In the case of the known family of myosin filaments not only are the myosin head arrangements under relaxing conditions being defined, but the latest analysis, also using single particle methods, is starting to reveal the way that the α-helical coiled-coil myosin rods are packed to give the filament backbones.

Keywords Muscle thick filaments • Titin • Tropomyosin • Myosin binding protein-C • C-protein • Myosin rod packing • Paramyosin • Troponin • Nebulin

11.1   Introduction and Overview of the Sarcomere

Myosin and actin filaments are most abundant in muscle, where they form the sarcomere in the striated muscles of vertebrates and invertebrates and the less ordered contractile units in vertebrate and molluscan smooth muscles (Squire 1981). After the discovery of the sliding filament mechanism in muscle by Huxley and Hanson (1953) and Huxley and Niedergerke (1953), which suggested that the sarcomere contains separate overlapping sets of thick and thin filaments (Fig. 11.1a middle; see also Squire (2016)), each of which remain more or less constant in length, a concentrated effort was made over the next few years towards defining the nature of these filaments. It was Hugh Huxley (1963) who first showed that the thick filaments contain myosin molecules, which are long rod-shaped molecules with a globular region at one end (Fig. 11.1a; top). He found that these molecules pack together with their rods antiparallel in the middle of a myosin filament and parallel towards each end, with the globular regions of the myosin molecules arrayed on the filament surface (Fig. 11.1a). Vertebrate muscle myosin filaments are bipolar; a 180° rotation around an axis perpendicular to the long axis of the filament and halfway along its length would leave the filament looking unchanged.

Fig. 11.1 (continued) end, aggregate to form the myosin filament backbone (*green*) with the myosin heads (*yellow*) arrayed on the filament surface in a helical or quasi-helical array. Actin filaments (*lower image*) contain actin monomers (*grey spheres*) together with strands of tropomyosin (blue lines) and troponin (*blue sphere*) (Adapted from Squire et al. 2005) (**b**) Summary of the biochemical cycle of adenosine triphosphate (ATP) hydrolysis by myosin (M) and actin (A) as proposed by Lymn and Taylor (1971). ATP hydrolysis occurs with myosin detached from actin (M.ATP to M.ADP.Pi). ADP is adenosine diphosphate. M.ADP.Pi attaches to actin and while attached Pi (inorganic phosphate) and then ADP are released during which at some point force and movement are generated. Further binding of ATP to the myosin heads causes head detachment from actin (Adapted from Squire et al. 2005)
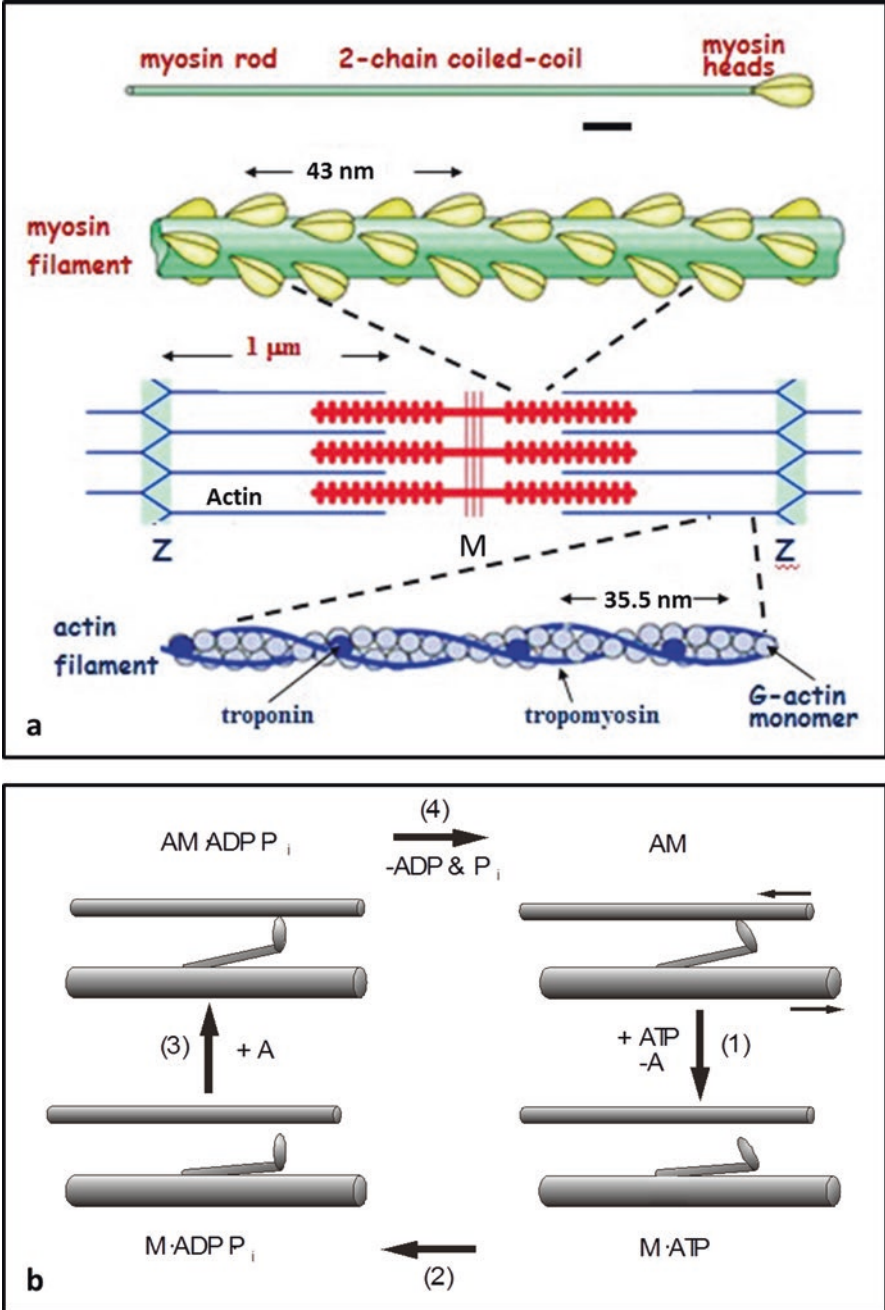
Fig. 11.1 (**a**) Schematic drawings of a striated muscle sarcomere (*central image*) showing overlapping myosin filaments (*red*) centred on the M-band (M) and actin filaments anchored to the Z-line (Z). Myosin rods (*top*), two-chain α–helical rods with two globular domains (heads) at one

The basic structure of actin filaments was first visualised by Hanson and Lowy (1963) using electron microscopy of isolated actin filaments viewed in negative stain. Actin filaments were seen to look like two strands of beads twisting around each other (Fig. 11.1a lower figure). The precise dimensions of these filaments were contentious, but by the time of the publication of the monumental X-ray diffraction study of frog *sartorius* muscle by Huxley and Brown (1967) the actin filaments were thought to be helical with a long repeat of around 35–36 nm and containing roughly 13 actin subunits in six turns of the genetic helix. In this same paper it was proposed incorrectly that the myosin heads on the myosin filaments in frog muscle were arranged on a 2-start helix, with a subunit axial translation of 14.3 nm (the characteristic axial spacing of all types of myosin filament) and a true repeat after about 43 nm. The spacings were correct, of course, but the symmetry proved to be wrong as explained later.

In the vertebrate striated muscle sarcomere (Fig. 11.1a; middle), the myosin filaments, cross-linked at their centres by M-band proteins, define the A-band (e.g. see Squire et al. 2005). They are overlapped at each end by actin filament arrays of opposite polarity, emanating from the Z-bands at the sarcomere edges. In cross-section, the myosin filaments form a hexagonal array, and in vertebrate striated muscles the actin filaments are located at the middle of a triangle formed by three adjacent myosin filaments (the so-called trigonal point).

It was also found early on that the globular part of myosin seen by Huxley was actually two globular domains, known as the myosin heads, on the end of a 150 nm long two-chain coiled-coil α–helical myosin rod (Slayter and Lowey 1967; Lowey et al. 1969; Fig. 11.1a top). These heads are ATPases, which can bind to actin, especially strongly in the absence of ATP, and the myosin head ATPase is activated by the interaction with actin. An early idea (e.g. Huxley 1969) was then that the myosin heads would bind reversibly to actin during muscle contraction in a cycle powered by ATP and that, when attached to actin, the heads would change their configuration on actin in some way, thereby producing movement (Fig. 11.1b; Huxley 1969; Lymn and Taylor 1971). But how was the muscle switched on? Early work had shown that tropomyosin, another two-chain coiled-coil α–helical protein (see Chap. 9), was associated with actin filaments. Then, in a ground-breaking study, Ebashi and his colleagues (Ebashi and Endo 1968) discovered the protein troponin on actin filaments in muscle, a protein which would reversibly bind calcium ions in the physiological range of calcium concentrations. In this scenario, an electrical stimulus travelling down the muscle nerve and into the t-tubular system would cause the release of calcium ions from the sarcoplasmic reticulum which would then bind to troponin on the actin filaments and, somehow, the actin filaments would be switched on.

The mechanism of switching was the subject of painstaking X-ray diffraction work by two groups, Hugh Huxley and his colleagues in Cambridge, England (Huxley 1972) and Peter Vibert and his colleagues in Aarhus, Denmark (Vibert et al. 1972). What they found was very striking. X-ray diffraction patterns from the actin filaments in muscle are characterised by so-called layer lines (Fig. 11.2, horizontal lines of intensity) with positions (spacings) that are orders of the basic actin filament axial repeat of around 35.5 nm. The first layer line corresponds to a repeat of 35.5 nm, the 2nd layer line to a spacing of 17.8 nm, the 3rd to a spacing of 11.8 nm and
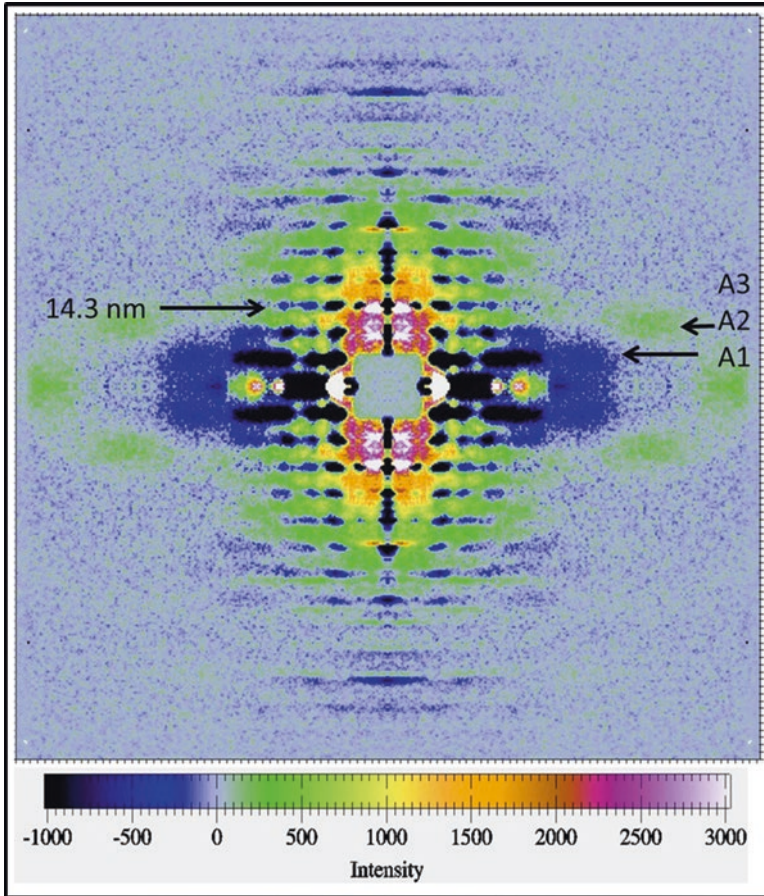
Fig. 11.2 Low-angle X-ray diffraction pattern from a vertebrate striated muscle (*fish muscle*) oriented with its long axis up the page and shown as a difference intensity map of the pattern from relaxed muscle subtracted from the pattern from fully active muscle (plateau of an isometric tetanic contraction). The key at the bottom shows how increases and decreases are depicted in the difference pattern. Of particular note is the increase in intensity (*green*) of the 2nd actin layer line (A2) in patterns from active muscle. The 3rd myosin layer line at 14.3 nm is also indicated (Modified from Mok 2005)

so on. These authors found that in X-ray diffraction patterns from resting muscle the 2nd layer line was relatively weaker than the 3rd, whereas in patterns from active muscle the 2nd layer line became much stronger, overtaking the 3rd which, in turn, became weaker. Figure 11.2 shows such changes in a more recent study and these are illustrated as a difference map showing which layer lines change (Mok 2005).

What did these changes mean? Ebashi had already suggested that tropomyosin molecules might lie along the grooves of the actin helix, with a troponin complex on each tropomyosin molecule. The tropomyosin molecules are about 40 nm long and would cover about seven actin subunits. We know that calcium ions would bind to
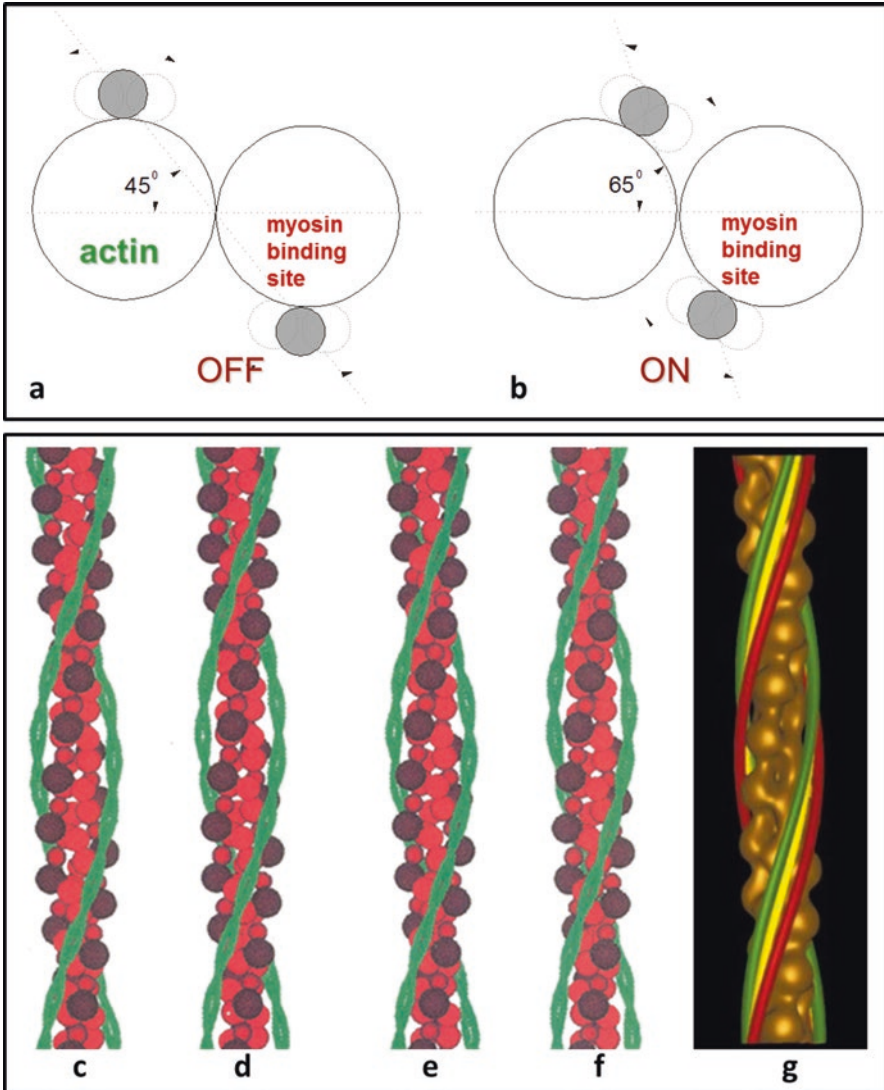
Fig. 11.3 (**a**, **b**) Illustration of the steric blocking model of muscle regulation as depicted in Parry and Squire (1973). The two strands of actin monomers are shown in cross-section as large circles. The small *grey circles* show the position of the tropomyosin strands under different conditions. In relaxed muscle (**a**, off, low calcium) the tropomyosins are well out of the grooves between the two actin strands. With elevated calcium levels (**b**), causing calcium binding to troponin, the tropomyosin strands move in towards the groove, thus uncovering the myosin binding sites on actin. The small *hollow circles* show the possible extent of small oscillations of the tropomyosin strands around their preferred location. It is as well to remember that the structure will never be static. Adapted and modified from Parry and Squire (1973). (**c** to **f**) Interleaved stereo pairs of a model

Fig. 11.4 3D reconstruction of actin filaments labelled with myosin heads in the absence of ATP showing the characteristic arrowhead structure and revealing where on actin the myosin heads bind
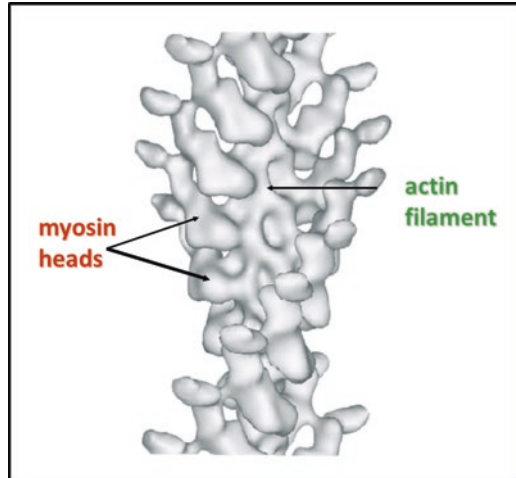
troponin in active muscle, but why would the layer lines change? The answer was shown at the same time by Huxley (1972), Haselgrove (1972) and Parry and Squire (1973). They found that tropomyosin strands were not actually right in the grooves of the actin filament in resting muscle, but were displaced laterally from the middle of the grooves (Fig. 11.3a, c, e). It was suggested that their position in resting muscle was such that the tropomyosin would block myosin head attachment, either directly or indirectly, so that contraction could not occur (Fig. 11.3a). As discussed later, the myosin head position on actin had been determined by electron microscopy and image processing of actin filaments labelled with myosin heads in the rigor state (Moore et al. 1970; a more recent structure is shown in Fig. 11.4).

When calcium ions became bound to troponin in active muscle, the effect seemed to be to move the tropomyosin strands closer to the grooves of the actin filaments, thereby exposing the myosin binding sites on actin (Fig. 11.3b, d, f). This mechanism became known as the 'Steric Blocking Mechanism'. All these authors showed that this movement probably occurred, but Parry and Squire (1973) also showed that every other type of movement of tropomyosin that they could think of failed to explain the observations; the tropomyosin shift appeared to be the only change possible. They also showed that tropomyosin molecules have a periodic repeat along them which divided the molecules into seven or 14 repeats (confirmed by Parry 1975; McLachlan and Stewart 1976), supporting an earlier conclusion from X-ray diffraction from tropomyosin tactoids (Caspar et al. 1969). The axial repeat of tro-

---

Fig. 11.3 (continued) actin filament showing tropomyosin strands (*green*) in the off position (**c** and **e**) and in the on position (**d** and **f**). The actin monomers are represented as four spheres in various shades of *red* (cf Fig. 11.5a). (**g**) Results based on 3D reconstructions of actin filaments in various states showing the three positions of tropomyosin associated with the off (*red*), calcium activated (*yellow*) and myosin head-bound (*green*) positions of tropomyosin ((**c–f**) from Squire and Morris 1998. (**g**) Reproduced from Craig and Lehman (2001) with permission)

ponin on the actin filaments was about 38.5 nm (Huxley and Brown 1967), so 38.5/7 = 5.5 nm coincided with the axial separation of actin monomers along one long pitched strand of actin; one tropomyosin molecule could make similar (not identical) interactions with seven actin monomers. The regulatory unit in the thin filament was, therefore, seven actin monomers, one tropomyosin molecule and one troponin complex, as originally suggested by Ebashi and Endo (1968).

With these basic ideas about myosin and actin filaments established, structural studies of both filament types since the early 1970s have progressed in gradually improving steps dictated by the available techniques. Here we briefly describe the methods used and the results that have been obtained.

11.2   Actin Filament Structure

11.2.1   3D Reconstruction of Actin Filaments

The first big step forward came from the realisation that electron micrograph images of extended helical objects, like actin filaments, lying down on the electron microscope grid actually contain many views of the same object, but at different rotations (i.e. viewing angles) around the long axis of the objects. We know that being able to view objects with our own two eyes means that we get a 3D impression of those objects. Therefore, a single image of a helical object in the electron microscope presenting many different views of the same object ought to contain the information necessary to reconstruct the object in 3D. Aaron Klug and his colleagues worked out how to do this using the computed Fourier transform of the object in an electron micrograph (DeRosier and Klug 1968), and this technique was put to good use on helical actin filaments by Moore et al. (1970), see also DeRosier and Moore (1971). The method was particularly useful when applied to actin filaments with myosin heads bound in the absence of ATP (Fig. 11.4). Here the curved, angled, elongated myosin heads were seen to produce a characteristic 'arrowhead' appearance, an appearance that has since been used to define actin filament polarity in a variety of studies. The reconstruction also showed where on actin the myosin heads appeared to bind, and it was this position that enabled the 'steric blocking model', mentioned earlier, to suggest that 'off' state tropomyosin was in the right place to block myosin head attachment.

11.2.2   Checking the Steric Blocking Model

The early X-ray diffraction evidence for the steric blocking model was itself fairly compelling. However, it is always good to be able to see structures directly. In a series of papers, Roger Craig, William Lehman, Peter Vibert and their colleagues

(Lehman et al. 1994, 1995; Vibert et al. 1997) studied actin / tropomyosin / troponin structures in a variety of states and from different muscles using the helical reconstruction method of Moore et al. (1970). Use was made of the roughly 13/6 helical symmetry of the actin filaments so that every actin monomer, together with its tropomyosin sub-repeat, was averaged together. It was known from the start that this was an approximation, since the seven tropomyosin sub-repeats in one molecule are not identical. But it was a good starting point and the result was that tropomyosin strands were seen in three different positions on actin (Fig. 11.3g); see Vibert et al. 1997 and many references therein). The three states were the 'off' state, thought to be where myosin head attachment was almost completely blocked, an intermediate state which resulted from calcium activation of the thin filaments, where the tropomyosin had shifted laterally as expected (about 20° rotation around the actin filament axis from the off state), and a third position, another 10° rotated, which resulted when myosin heads were bound in a strong state (e.g. rigor, AM; Fig. 11.1b). These three states were identified by Lehrer and Morris (1982) with the "off", "on" and "potentiated" states characterised by Bremel and Weber (1972). Further work substantiated these states (Phillips et al. 1986; McKillop and Geeves 1993, AL-Khayat et al. AL-Khayat et al. 1995; Brown and Cohen 2005; Poole et al. 2006). The states became known as the Blocked or 'B' state, the closed or 'C' state, where a limited amount of head binding could occur, and the fully open Myosin or 'M' state. Kinetic data were consistent with the tropomyosin being in different equilibria between the states, depending on the conditions (McKillop and Geeves 1993; Lehrer and Geeves 1998).

11.2.3   X-Ray Crystallography of Actin and Myosin

Actin filaments occur not just in muscle, but are fundamentally important in all cell types, so when Kabsch et al. (1990) solved the structure of the globular actin (G-actin) monomer by X-ray crystallography, the muscle and cell motility fields were advanced in a spectacular way. The G-actin monomer was found to have four domains in two pairs with a cleft between them, with ATP bound in the bottom of the cleft (Fig. 11.5a). With this structure to hand, and using high angle X-ray diffraction data from oriented gels of actin filaments, Holmes et al. (1990) were then able to suggest how these G-actin molecules might aggregate to form an actin filament (Fig. 11.5c). In brief, subdomains 3 and 4 are closest to the actin filament axis, where they interact with actin monomers in the other strand of the filament, and subdomains 1 (large) and 2 (small) lie on the outside of the filament. Subdomain 1 provides the major binding site for myosin heads. The earlier helical reconstruction work probing the location of tropomyosin in different states also put tropomyosin in the 'off' state binding to subdomain 1 of actin where the heads want to attach.

    The field took another major leap forward when the myosin head was crystallised for the first time and its structure determined using X-ray crystallography by

Fig. 11.5 (**a**) Actin monomer with each of its four structural subdomains shown in different colours, subdomain 1: *red*, subdomain 2: *green*, subdomain 3: *blue*, and subdomain 4: *yellow*. (**b**) Schematic of the whole troponin complex with a globular region composed of Tn-C, Tn-I and part of Tn-T (Tn-T2: residues 159-259) with a rod region of Tn-T (Tn-T1: residues 1-158). The crystal structure of Tn-C is shown (Herzberg and James 1988). (**c**) The helical arrangement of the actin monomers along the actin filament, F-actin, according to Holmes et al. (1990). (**d**) Tropomyosin molecules consist of a two-chain α-helical coiled coil (Diagram modified from Squire and Morris 1998)
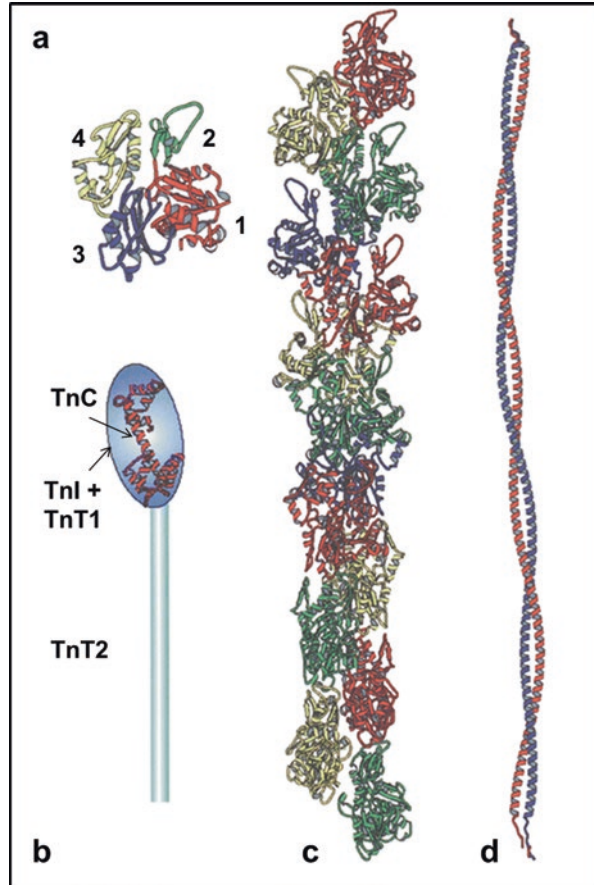
Rayment et al. (1993a, b). This showed the head to be a remarkably interesting structure (Fig. 11.6a). It consists of a motor domain, which contains the ATPase site, and the part of the head that binds to actin. This is then linked through a so-called converter domain to a long α-helical region around which wrap two myosin light chains, the essential light chain (ELC) and the regulatory light chain (RLC). This region is called the tail or lever arm of the myosin head. It was then found that, with different ATP analogues bound to the heads (Rayment et al. 1993b; Dominguez et al. 1998; Houdusse et al. 2000), the motor domain remained fairly constant in structure (apart from a cleft opening and closing; Holmes et al. 2004), but that the angle, which the lever arm made with the motor domain, could vary significantly. This immediately gave rise to the idea that the motor domain might bind to actin in a fairly well defined way, and that muscular movement could be associated with the lever arm swinging on the relatively fixed actin-attached motor domain, thus tending to make the actin and myosin filaments slide past each other.
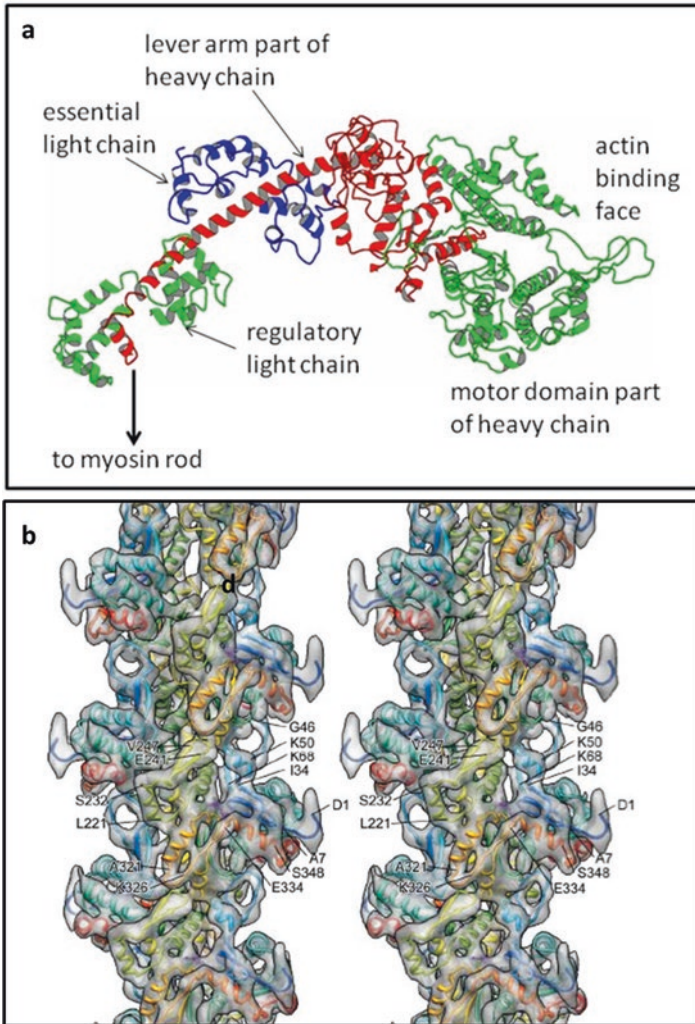
Fig. 11.6 (**a**) Structure of the S1 fragment of myosin (the myosin head) showing the heavy chain (*right green/red* part) and the α–helical region (*red*, *left*). The essential light chain is shown in *blue* and the regulatory light chain in *green* (*left*). (**b**) Good resolution (0.66 nm) structure of rabbit skeletal muscle F-actin as determined by Fujii et al. (2010; central part of their Fig. 11.2) using cryo-electron microscopy. Some amino acids are highlighted. ((**b**) Reproduced by permission Fujii et al. 2010)

11.2.4 Recent High Resolution Thin Filament Helical Reconstructions

The early helical reconstructions of actin alone or with tropomyosin/ troponin and heads bound were usually carried out on filaments that had been negatively stained for electron microscopy. However, the electron microscopy of frozen filaments,

although much more difficult, allows significantly higher resolutions to be achieved (Xu et al. 1999). In addition, methods of analysis of electron microscope images of helical filaments have been much improved (e.g. Egelman 2000). Here the helical filaments can be divided up into short 'particles', which are analysed and averaged in a procedure which optimises the estimated helical parameters. These are then used in calculating the final reconstruction. The use of relatively short filament segments minimises the degradation of the reconstruction by deviations from pure helical symmetry that may occur along a long piece of filament (Li et al. 2011).

A relatively recent result from this approach applied to actin filaments is the 0.66 nm resolution structure determined by Fujii et al. (2010) and displayed in Fig. 11.6b. Here the secondary structures in the actin molecules (α-helices, β-sheets, loops etc) can be seen clearly. The known crystal structure of G-actin (Otterbein et al. 2001; Protein Data Bank ID, 1J6Z) could be fitted readily into the density map and it could then be refined using simulated annealing molecular dynamics with stereochemical and non-bonded interaction terms restrained (Flex-EM; Topf et al. 2008). The new structure showed that interactions along the filament were relatively strong, but that those between the two long-pitched strands of actin were comparatively loose. A more recent reconstruction of actin-tropomyosin at 0.37 nm resolution for F-actin and 0.65 nm resolution for tropomyosin has also been published (von der Ecken et al. 2015).

Another significant advance was achieved when the structure of actin-tropomyosin-myosin in the rigor state was determined using similar methods at around 0.8 nm resolution (Behrmann et al. 2012). This structure (Fig. 11.7) reveals interactions between a single myosin head and two successive actin monomers along a long-pitched strand, together with one tropomyosin sub-repeat. As expected the tropomyosin is positioned in the 'M' state, about 23° rotated about the thin filament axis from the 'off' (closed or blocked; B) position. This was not quite the 30° of Vibert et al. (1997), but was more rotated than their "C" state. What is very apparent is that the rigor myosin heads and the tropomyosin strands are tightly interacting.

One of the problems of the helical reconstruction approach applied to thin filaments, when trying to probe the location of tropomyosin in different states, is that troponin is still present (Vibert et al. 1997). Troponin is only on every seventh actin monomer, so when helical averaging takes place the troponin density is smeared out across seven actin monomers instead of one (as detailed by Squire and Morris 1998). Information about troponin is lost and the density that appears to be tropomyosin actually has troponin density added into it. Since the structure of troponin presumably changes when it binds calcium, some of the observed apparent shift of tropomyosin could be an artefact due to troponin movements and shape changes. In short, another technique is needed to reveal troponin and to avoid the artefacts of helical averaging. This is where true single particle analysis comes in.

Fig. 11.7 Good resolution (0.8 nm) structure of actin plus tropomyosin plus myosin subfragment 1 as determined by Behrmann et al. (2012). F-actin was prepared from rabbit skeletal muscle, and tropomyosin and myosin (MyoE) were purified from *Escherichia coli* (E. coli) and *Dictyostelium discoidium* cells respectively. Pseudo atomic models of the proteins have been fitted to the calculated density map (Reproduced by permission of Behrmann et al. 2012)
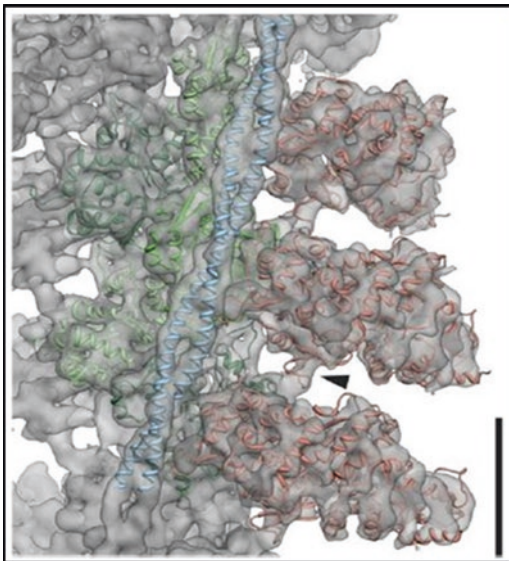
11.2.5 Finding Troponin – Single Particle Analysis

The method of single particle analysis was developed late last century to generate 3D reconstructions of roughly spherical objects such as ribosomes and viruses (e.g. van Heel et al. 1996, 2000; Frank et al. 1996; Ludtke et al. 1999). The technique recognised that fields of globular particles lying on an electron microscope grid in negative stain, or preferably in a thin layer of ice, would probably lie in all orientations on the grid and would therefore provide multiple projected density views of the same object. As in tomography, the different views could then be recombined to produce a 3D model of the object. The viewing angle for each projection image (or class average formed from the average of many images corresponding to the same view) was then determined and the structure regenerated to give the required 3D density map. Spectacular results have been achieved for ribosomes and other protein complexes, exploiting recent developments in detector technology (Kuhlbrandt 2014; see also Chap. 1, Sect. 1.5).

Some while ago it was realised that even highly elongated particles, which show some form of axial periodicity, could be treated as single particles (e.g. AL-Khayat et al. 2004; Paul et al. 2004). In this case, images of long thin filaments could be chopped up into segments centred on the troponin density and each segment could then be treated as an image of a single particle. Different parts of the same filament or equivalent parts of different filaments would present different views of the same structure but rotated around the filament axis. Class averages could then be generated, the viewing angles found and reconstruction by back projection carried out.

This technique applied to thin filaments is significantly more difficult than helical reconstruction, mainly because each actin filament necessarily provides only
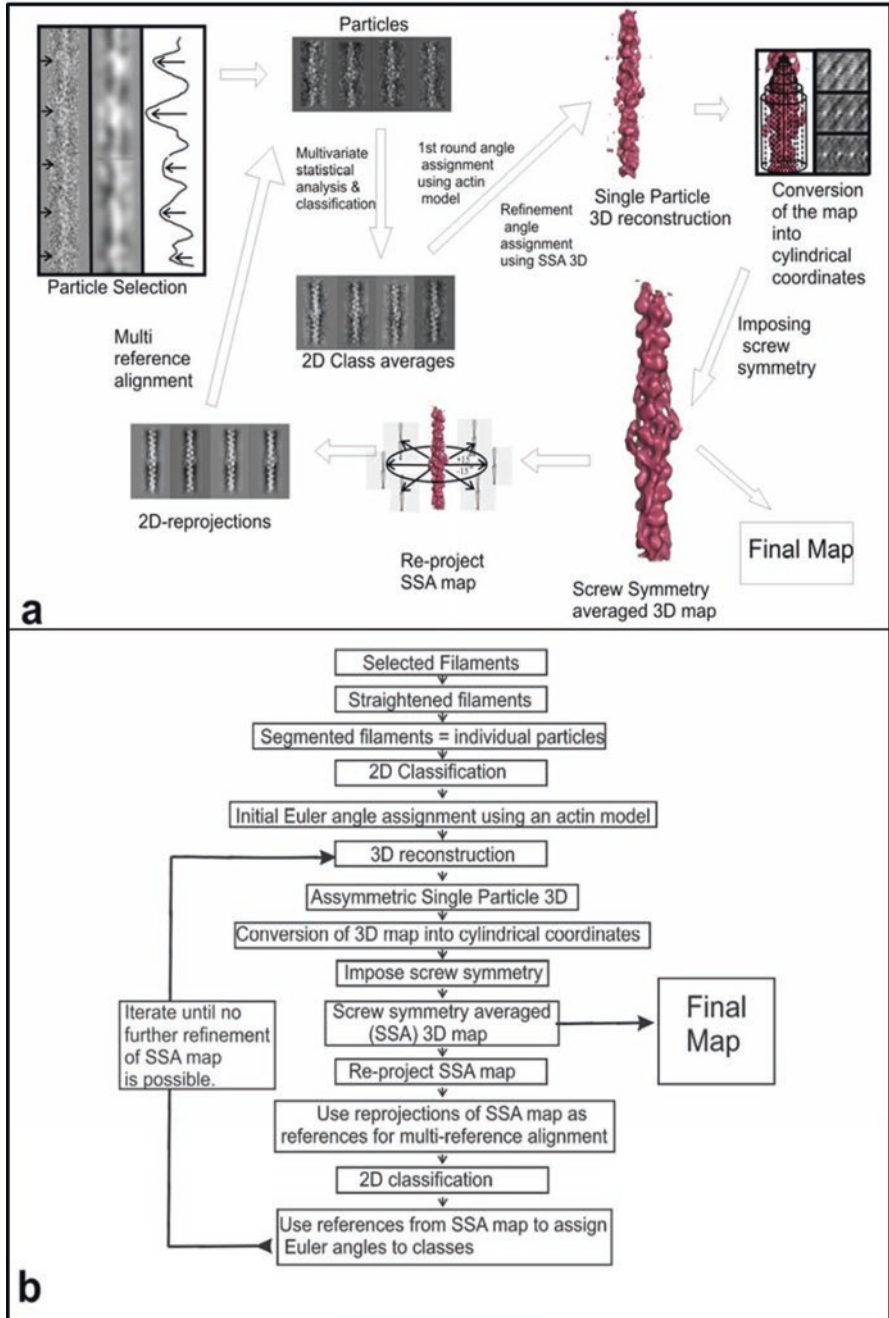
Fig. 11.8 Thin filament single particle image processing flow charts in the method of Paul et al. (2017). (**a**) The cyclical nature of the image processing steps is depicted. Included here is an example of the particle selection procedure where the filaments were visually inspected and the

one-seventh of the data that helical reconstruction has available. However, promising 3D reconstructions have already been obtained and the results are striking, even though the work used negatively-stained filaments and the resolution achieved is as yet only about 2.5 nm. The methods used by Paul et al. (2017) are summarised in Fig. 11.8. The most important step is to find the viewing angles of the different images. What is usually done is to compare the experimental images with projections of a reference structure seen at known viewing angles (projection matching). In the case of the thin filament, this has been done in one case by using as a reference object a hypothetical model of actin / tropomyosin / troponin (Pirani et al. 2005). However, the result is a 3D reconstruction that is very similar to the original reference structure with rather little extra density features, so it is likely that the reconstruction was not refined away from the starting structure. A better approach, when trying to discover the location of tropomyosin and the structure of troponin, is to use only F-actin as the reference. In this case there can be no bias in what is found out for tropomyosin and troponin, since they are not present in the reference structure. The results from this latter approach applied to thin filaments with and without calcium bound (Paul et al. 2017) are shown in Figs. 11.9 and 11.10. The tropomyosin in the 'off' state (low calcium; Fig. 11.9e, f) is in much the same position on all seven actin monomers in the repeating unit, as in the case of the helical reconstruction. The troponin density wraps around the filament on one actin monomer and also has a long tail of density, which extends along the filament (upwards in Fig. 11.9). The structure of the whole of troponin is not known from X-ray crystallography, but some parts of it have been solved including troponin-C plus parts of troponin I and troponin-T (Takeda et al. 2003; Vinogradova et al. 2005). These incomplete structures can be fitted quite well into the density in the 3D reconstruction. The remaining parts of troponin have been generated by using model building (Manning et al. 2011), and this arrangement, modified slightly, seems to fit quite well with what is seen (Fig. 11.10). It is known that TnC, together with TnI and part of Tn-T (TnT2; residues 159–259; Takeda et al. 2003; Vinogradova et al. 2005), form a globular complex and that the rest of Tn-T (Tn-T1; residues 1–158) probably extends out from the globular region as long α–helical strands. It is likely that it is this elongated

---

Fig. 11.8 (continued) identification of troponin density was guided by the use of axial density profiles and the appearance of Fourier-filtered images of the filaments in which the spatial frequencies ≤200Å and ≥500Å are suppressed. The axial density profiles from these filtered images showed peaks originating from the troponin density. (**b**) A description of the image processing cycle used in the analysis of both the +Ca^{2+} and –Ca^{2+} data. After segmenting the filament and generation of 2D class averages an actin model was used to assign Euler (viewing) angles to create the initial single particle 3D reconstruction. This asymmetric first 3D map was then converted into cylindrical coordinates and the screw symmetry of the two actin strands was imposed. Re-projections of this screw symmetry averaged 3D map were then used in a multi-reference alignment. The cycle of alignment, classification, 3D reconstruction, imposing screw symmetry and re-projection was repeated until no further improvement in the maps resolution was achieved. The final map was a screw symmetry-averaged single particle reconstruction
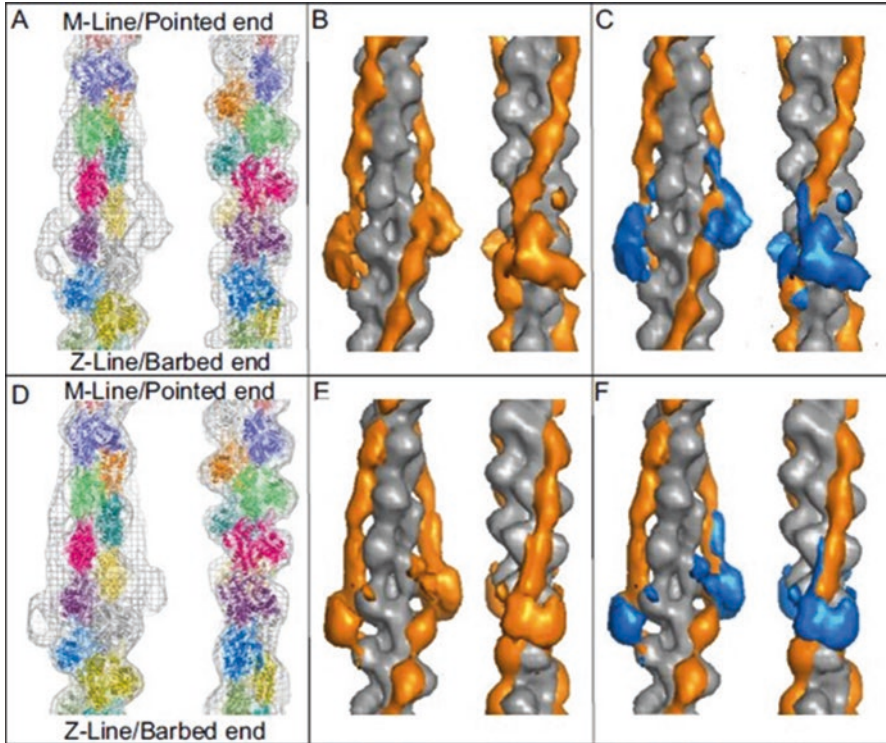
Fig. 11.9 (**a–c**) Data from Ca^{2+} -treated filaments, (**d–f**) data from Ca^{2+} -free filaments. (**a**) and (**d**) Wire mesh representations of the single particle-based reconstructions of the thin filament in the two states. An atomic F-actin model is docked into the reconstruction and each subunit is colour coded. The barbed end (Z-band end) of the actin filament is at the bottom of the figure. (**b**) and (**e**) Difference density maps calculated by subtracting the docked F- actin model (*grey*) from the single particle reconstructions. This leaves the density attributable to the regulatory proteins troponin and tropomyosin (both *orange*). (**c** and **f**) Difference density maps calculated by subtracting docked F-actin (*grey*) and tropomyosin (*orange*) models from the single particle reconstructions leaving density attributable only to troponin (*blue*). (Reproduced from Paul et al. 2017 with permission)

part of Tn-T that accounts for the finger-like extension in the troponin density that runs up the filament in Fig. 11.9c, f. One interesting feature of this is that the direction of the Tn-T tail is opposite to that reported earlier by Ohtsuki (1979). The Tn-T tail really does appear to be directed towards the M-band or pointed end of the actin filament, not the other way.

The results from the 3D reconstruction of the thin filament in the presence of calcium are quite unexpected. It has been commonly assumed that tropomyosin moves across the actin filament in much the same way on each actin, as in Fig. 11.3, even though the tropomyosin sequence repeats are only quasi-equivalent. But the actual structure is not like that. The tropomyosin repeats close to troponin on the pointed end side shift across the filament by about 18° (tropomyosin repeats d to g in Fig. 11.10b, which we refer to as set 1), but those on the other side of troponin
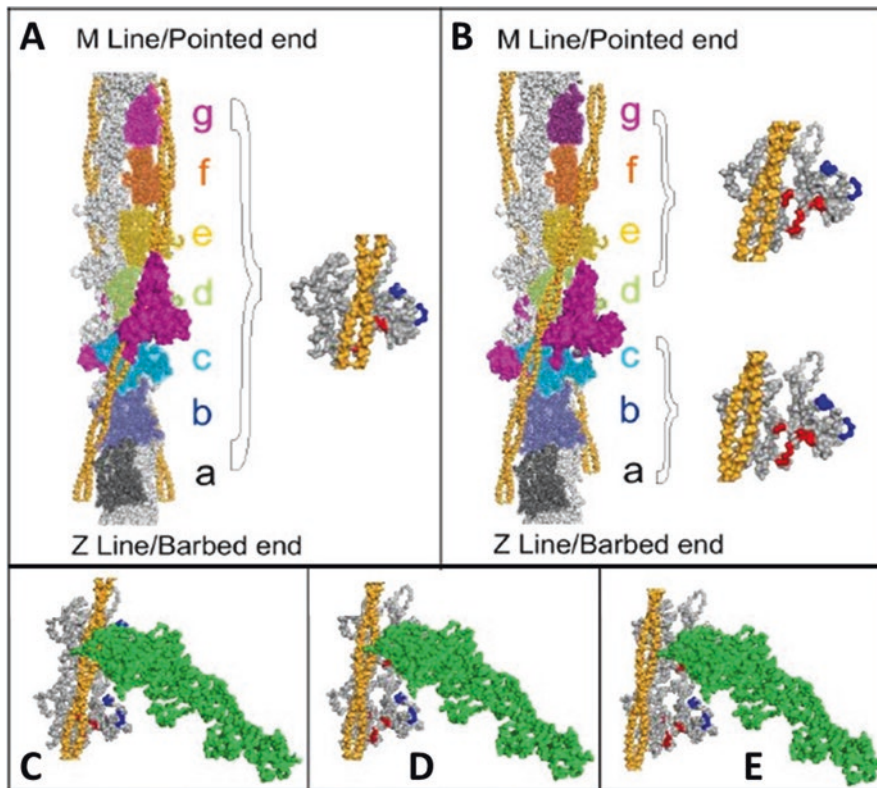
Fig. 11.10 Visualisation of the tropomyosin domain movements and myosin binding. (**a** and **c**) Structures of the Ca^{2+} -free state. (**b**, **d** and **e**) Structures of the Ca^{2+} -treated thin filament assembly. (**a**) and (**b**) The central region of the Ca^{2+} -free (**a**) and the Ca^{2+} -treated (**b**) reconstructions over a length of 14 actin subunits, two strands of tropomyosin (*orange*) and two troponin complexes. The core domains of the troponin are modelled using COMPHI, an atomistic model of cardiac troponin in the Ca^{2+}-treated state published by Manning et al. (2011). The model was generated using the existing crystal structure of the 52 kDa domain of human cardiac troponin in the Ca^{2+} -treated state (Takeda et al. 2003) together with missing regions that were constructed using secondary structure prediction. Subunits a to g of one strand of actin are colour coded and the second strand is *grey*. In the Ca^{2+}-free state (**a** and **c**) the position of tropomyosin is the same on every subunit and lies in the 'B' or blocked position where it would inhibit myosin binding. The position of tropomyosin is different on different subunits of the actin filament in the Ca^{2+} -treated state (**b**), where the average positions are illustrated (**a** and **b**; right inserts) for actin subunits d to g (closed state; set 1) and a to c (M-state; set 2). **c**, **d** and **e** show face on views of the actin subunits, with weak (*blue*) and strong (*red*) actin binding sites highlighted for the three distinct positions of tropomyosin. Subunits a, b and c show tropomyosin in the "M or Myosin state", subunits d, e, f, g show tropomyosin in a position more closely aligned to the "C or closed" state. In the Ca^{2+} -treated thin filament a myosin S1 head (*green*) can access all weak and potentially some strong binding sites on subunits e, f and g (**d**) and presumably all available binding sites on subunits a and b (**e**). An equivalent myosin head would be blocked on every subunit in the Ca^{2+} -free state (**c**) (Reproduced from Paul et al. 2017 with permission)

move on average by about 28° (tropomyosin repeats a to c to in Fig. 11.10b; set 2). So, in the presence of calcium, set 1 tropomyosin are almost in the old closed state of Vibert et al. (1997), but even in the absence of myosin, set 2 tropomyosin repeats are in what is close to the 'M' state. Rather than tropomyosin moving as a fairly rigid body, as part of a so-called Gestalt mechanism (Holmes and Lehman 2008), the tropomyosin shift varies from actin to actin so that some myosin binding sites are partially blocked and others appear to be completely open; the 'C' state of Squire and Morris (1998) appears not to be an equilibrium between C and M states, as in one of their original suggestions (Fig. 11.11a), but a defined mixture of 'C' and 'M' tropomyosin sub-repeats within the same tropomyosin molecule (Fig. 11.11b).

11.2.6   *The Thin Filament Regulation Mechanism*

The new varying 'on' position for tropomyosin in the presence of calcium, suggests a new scenario for thin filament activation in muscle. In the 'off' state the tropomyosin is presumably in a position where myosin head attachment is unlikely. But it is as well to remember that none of these proteins will be static and the occasional excursion of tropomyosin from its fully off position, which would allow occasional head binding, would not be surprising. In the calcium-activated filament, which we now know is a mixture of closed ('C') and myosin ('M') states, it is likely that myosin heads opposite some points along actin (near tropomyosin set 2) will bind more easily than those in other places (set 1) and limited contraction will proceed (Fig. 11.11b). The more heads that then bind in strong states, the more the whole of tropomyosin will be constrained to the 'M' state and eventually full activation will occur (Fig. 11.11d).

11.2.7   *Nebulin*

Chapter 10 in this volume discusses in detail another component of the thin filaments in vertebrate striated muscles. This is the giant protein nebulin. It has a chain weight of around 600–900 kDa and runs along the thin filaments in the I-band from the Z-band to the actin filament tip (Fig. 11.12a). Nebulin has never been visualised on the thin filament, but it is long enough to contribute a long mainly α–helical strand along the filament, possibly running together with tropomyosin. There are repeating units of about 35 residues which with a 0.15 nm subunit axial translation in an α–helix gives an axial repeat of 5.25 nm, close to the actin monomer separation in a long-pitch actin strand of 5.5 nm. These repeats have been shown to bind strongly to actin and they occur in groups of seven, corresponding to the length of the regulatory unit in the thin filament. It is not known how many nebulin strands
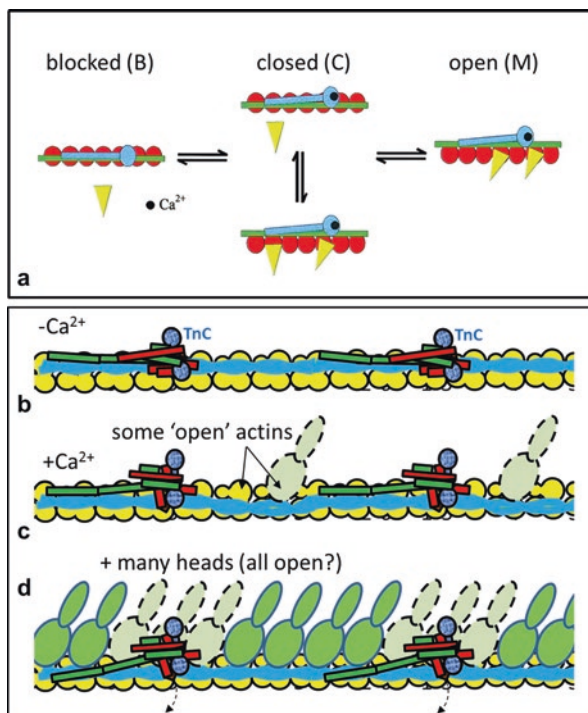
Fig. 11.11 (**a**) One of the possible mechanisms illustrated by Squire and Morris (1998) to explain the closed (**c**) state of thin filament regulation as being an equilibrium between the blocked and open states. (**b** to **d**) A novel mechanism of regulation unlike that in (**a**) where the 'C-state' is actually a mixture of the C and M states**.** Schematic diagrams of the thin filament assembly for simplicity are shown as a linear array (i.e. not helical). Protein components colour coded: actin (*yellow*), tropomyosin (*blue*) and troponin (TnC, speckled *blue*; TnI, *red*; TnT *green*). When the thin filament changes from the Ca^{2+} -free state (**b**) to the Ca^{2+}- bound state (**c**) the tropomyosin strands on the arrowed central actins (set 2) move further than those on actins adjacent to the Tn-T1 tail (*green*; set 1). Set 2 actins may be exposed enough for some myosin heads to bind (*dashed* outline). Bound heads going over to strong states may then activate the whole filament (**d**). The body of troponin may move out of the way (*dashed* arrows) to permit binding of further heads. (**b** and **c**) also illustrate a potential mechanism whereby the distal arm of TnT1 acting on the tropomyosin overlap region may swing to expose the binding sites on set 2 actins when Tn-C binds Ca^{2+} ((**a**) (Reproduced with permission from Squire and Morris (1998). (**b–d**) reproduced with permission from Paul et al. (2017))

there are per thin filament, but it is quite likely to be two. Nebulin occurs in many splice isoforms of varying length and this length has been shown to determine the length of the actin filaments in the I-band. The thin filament is terminated at the end toward the M-band (the pointed end) by the protein tropomodulin (McElhinny et al. 2001). Other proteins associated with nebulin are desmin, titin and capZ (Fig. 11.12b) as discussed in Chap. 10 of this volume.

Nebulin is not just a structural protein. It has been shown that its presence on the thin filaments enhances the level of force produced by the sarcomere (Fig. 11.12c),
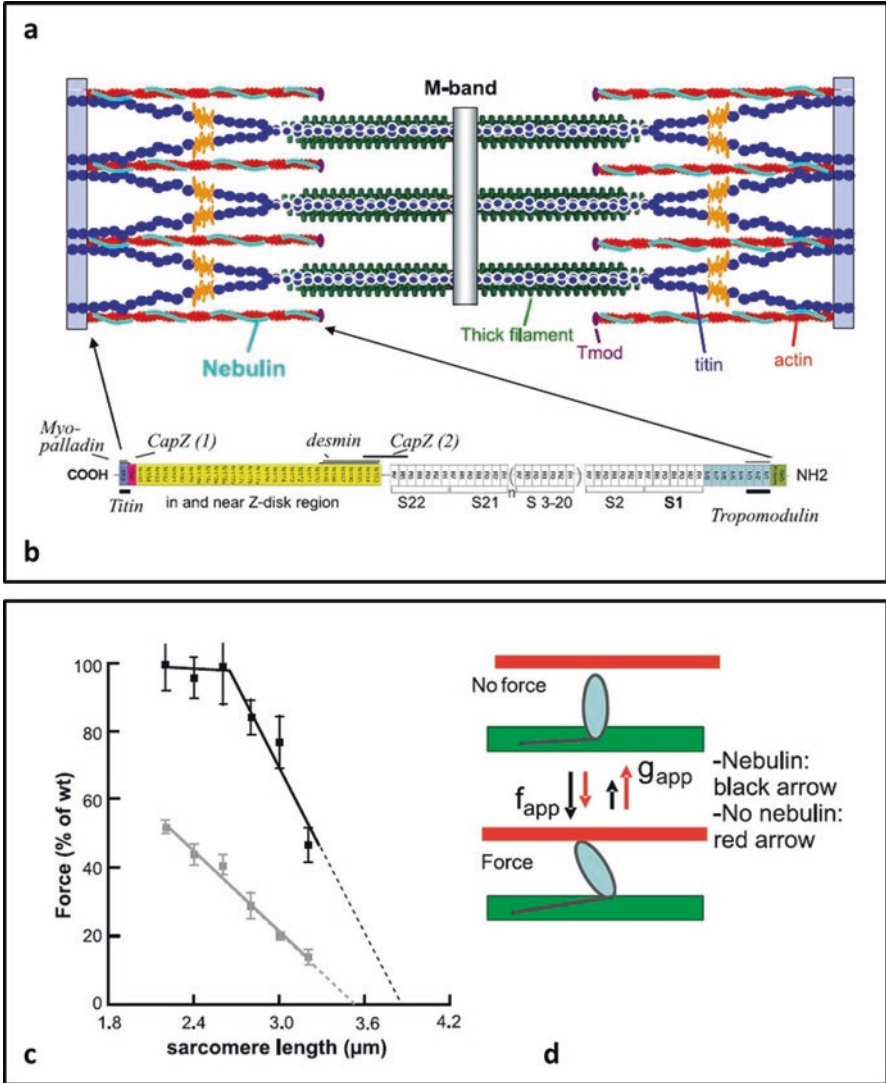
Fig. 11.12 Illustrations of the location and behaviour of nebulin on muscle thin filaments. (**a**) Layout of nebulin in the skeletal muscle sarcomere. Titin strands are also shown. (**b**) Modular organization of nebulin protein. Binding sites of nebulin-binding proteins are also shown: myopalladin, titin, the two reported capZ binding sites, CapZ(1) and CapZ(2), desmin and tropomodulin. (**c**) Comparison of the measured force sarcomere length relation of wild type and nebulin deficient NEB-KO muscle fibres. (**d**) Illustration of a 2-state crossbridge cycle and the effect of nebulin on the rate constant of transition from non-force-generating cross-bridges to force-generating cross-bridges f_{app} and on the reverse rate constant g_{app}. Nebulin increases f_{app} and slows g_{app}, which has the net effect that the fraction of all available heads that generate force is increased (Reproduced from Labeit et al. 2011 with permission)

possibly changing the attachment and detachment rates of the myosin heads on actin (f_{app} and g_{app}; Fig. 11.12d). Mutations of nebulin have also been shown to be associated with various myopathies (e.g. Ottenheijm et al. 2013 and references therein and Chap. 10).

The ease with which myosin heads can bind to actin also depends on where they are coming from on the neighbouring myosin filaments. This is the topic of the rest of this Chapter.

11.3 Myosin Filament Structure

11.3.1 Myosin Filament Symmetries

Vertebrate striated muscle myosin filaments are bipolar, with centrally packed anti-parallel myosin molecules and outer parallel myosin molecules (Fig. 11.13a, b). The filaments therefore have a central bare zone and outer regions where the myosin heads project – the bridge regions (Fig. 11.13b). The first idea that the myosin heads on myosin filaments might be relatively regularly organised on the myosin filament surface came from the pioneering electron microscopy by Huxley (1963), and the X-ray diffraction studies of frog muscle by Huxley and Brown (1967) and Elliott (1964a; b, 1967). Huxley and Brown (1967) proposed that the myosin heads in frog muscles were organised on 2-start helices with an axial shift between successive subunits along the helices (the subunit axial translation) of 14.3 nm. The whole structure would repeat after 43 nm. Shortly after this, in a pioneering study of the appearances of longitudinal sections of insect flight muscle (IFM), Reedy (1968) proposed that the IFM myosin filaments, with a bigger backbone diameter than vertebrate filaments, were also on a 2-start helix with a subunit axial translation of 14.5 nm, but with a long repeat of 115 nm. In 1972, Small and Squire (1972) published a structure for the face polar myosin ribbons seen in vertebrate smooth muscle under some conditions (see Sect. 11.3.9), where the heads also had an axial shift of 14.3 nm, but this time there was an apparent long repeat of around 72 nm. Remembering that the myosin molecules making up these filaments were all rather similar in structure (a long two-chain coiled-coil α–helical rod with two heads at one end (Fig. 11.1a top), it seemed to Squire (1971, 1972) that such similar molecules would not pack in such different ways. He suggested that vertebrate filaments might have their heads on 3-start or 4–start helices and that IFM filaments might have up to six coaxial helical strands. In this way the molecular packing of the myosin rods in these different thick filaments would all be rather similar. In the end it was found that vertebrate myosin filaments had their heads on 3-start helices (Squire 1972, 1973, 1981, 1986; Maw and Rowe 1980; Kensler and Stewart 1983) and IFM filaments on 4-start helices (Wray 1979; Reedy et al. 1981; Morris et al. 1991). Later work also showed that the heads on the filaments in scallop striated muscle were on 7-start helical strands (Millman and Bennett 1976; Vibert and Craig 1983; Craig et al. 1991; Vibert 1992; AL-Khayat et al. 2009a, b; Woodhead et al. 2013).
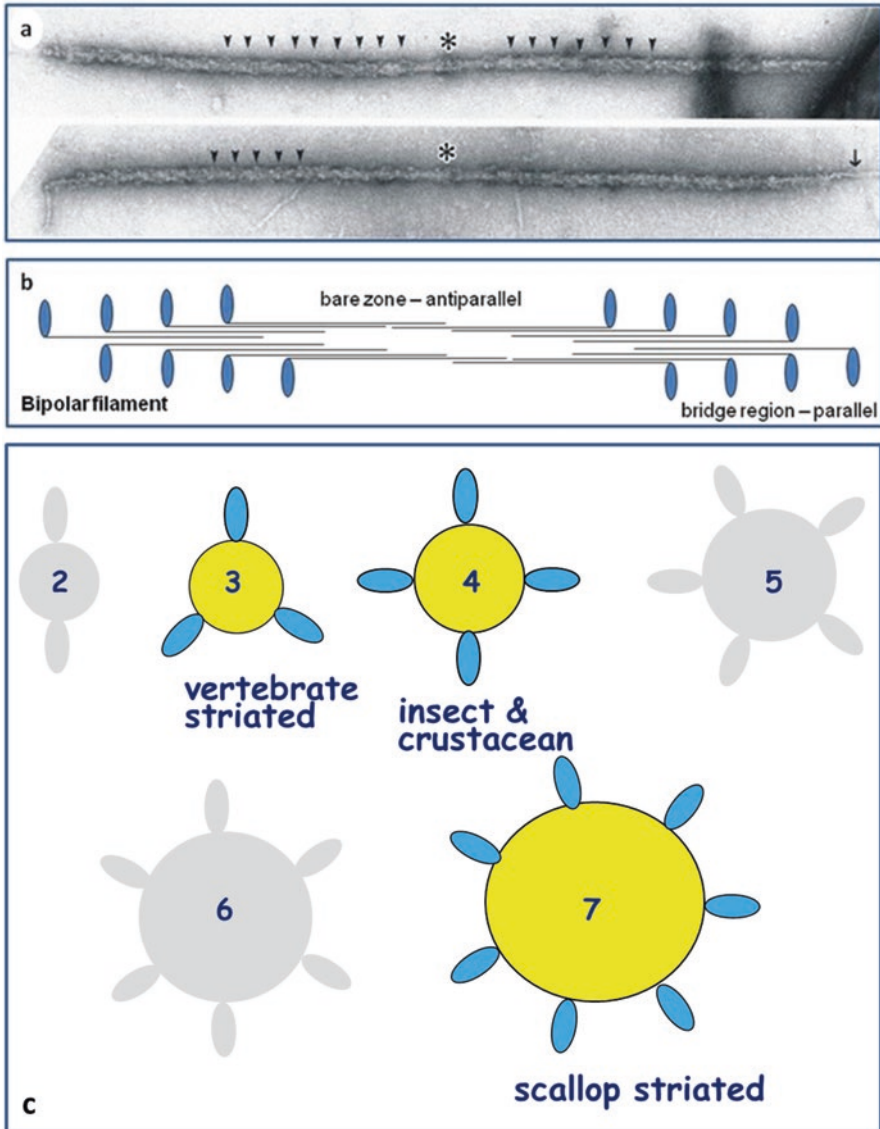
Fig. 11.13 (**a**) Electron micrograph of isolated bipolar thick filaments from fish muscle showing the axial repeat patterns on each side of the central M-band region (starred). (**b**) Schematic diagram to show the antiparallel arrangement of myosin molecules in the bare zone of striated muscle thick filaments and the parallel packing in the filament bridge regions. The bridge regions are very much longer in real filaments. (**c**) Illustration of the possible rotational symmetries that myosin filaments might possess. Those filaments identified so far are the 3-start, 4-start and 7-start helices. 2-start filaments are probably physically impossible, but it is conceivable that 5-start or 6-start filaments might be found

All these structures are illustrated in Fig. 11.14, where it can be seen that the axial and lateral separation of the myosin heads on the filament surfaces are all rather similar. Figure 11.15 compares the low-angle X-ray diffraction patterns from fish muscle (a) and insect flight muscle (b), both of which show superb 3D order, similar meridional peaks at around 14.3–14.5 nm, but different layer line spacings revealing the different myosin head lattice geometries.

Hanson and Lowy (1963), Szent-Gyorgyi et al. (1971) and Hardwicke and Hanson (1971) showed that the very large myosin-containing filaments in molluscan smooth muscles have a thin surface layer of myosin on a core of paramyosin. Paramyosin molecules are rather like the rod part of myosin without the heads at one end. Squire (1971) showed that the amounts of myosin and paramyosin in muscles with different filament diameters were consistent with the myosin packing into similar 2D nets of heads to those in Fig. 11.14a–d, with a central core of paramyosin molecules. The structures of these paramyosin filaments will be discussed later in the Chapter (Sect. 11.3.8).

Figures 11.13c and 11.14a–d show that the known myosin filaments form a family with different rotational symmetries – 3-start (vertebrate), 4-start (IFM, *Limulus*, crab etc) and 7-start (scallop). Apart from very large paramyosin filaments, are these the only possibilities? For example, could there be 5-start and 6-start filaments in some muscles, probably invertebrate, that have not yet been discovered (Fig. 11.13c)? In a detailed analysis of the X-ray diffraction patterns from some crustacean muscles, Wray (1979) suggested that myosin filaments in lobster abdominal flexor muscle might be 4-stranded, whereas those from the lobster crusher claw muscle might be 5-stranded, but this has yet to be confirmed.

11.3.2 Myosin Head Organisation

With many myosin filament lattice symmetries known, the next question to ask is how are the myosin heads arranged on these surface lattices? Fig. 11.14e–g illustrate schematically the regular ways in which myosin heads might conceivably be organised if the heads are stabilised by interaction with other heads. Each layer of heads spaced by 14.3–14.5 nm is usually known as a 'crown' of heads. The two heads of a single molecule might interact with each other and might point up the same long period helices as in the type 1 configuration (Fig. 11.14e). Alternatively they might interact across the filament with heads from adjacent molecules in the same crown (type 2; Fig. 11.14f) or they might interact with heads from other molecules in successive crowns along the filament (type 3; Fig. 11.14g). Early analysis of X-ray diffraction data suggested that the elongated heads in vertebrate striated muscles might be tilted up the filament, perhaps along the long period helices as in the type 1 structure, to explain why the 1st and 2nd (42.9 and 21.5 nm) layer lines in X-ray diffraction patterns (Fig. 11.15a) were relatively strong compared with the meridional M3 reflection at 14.3 nm (Squire 1975). The structures of isolated myosin filaments (e.g. Fig. 11.13a) were then studied using electron microscopy and
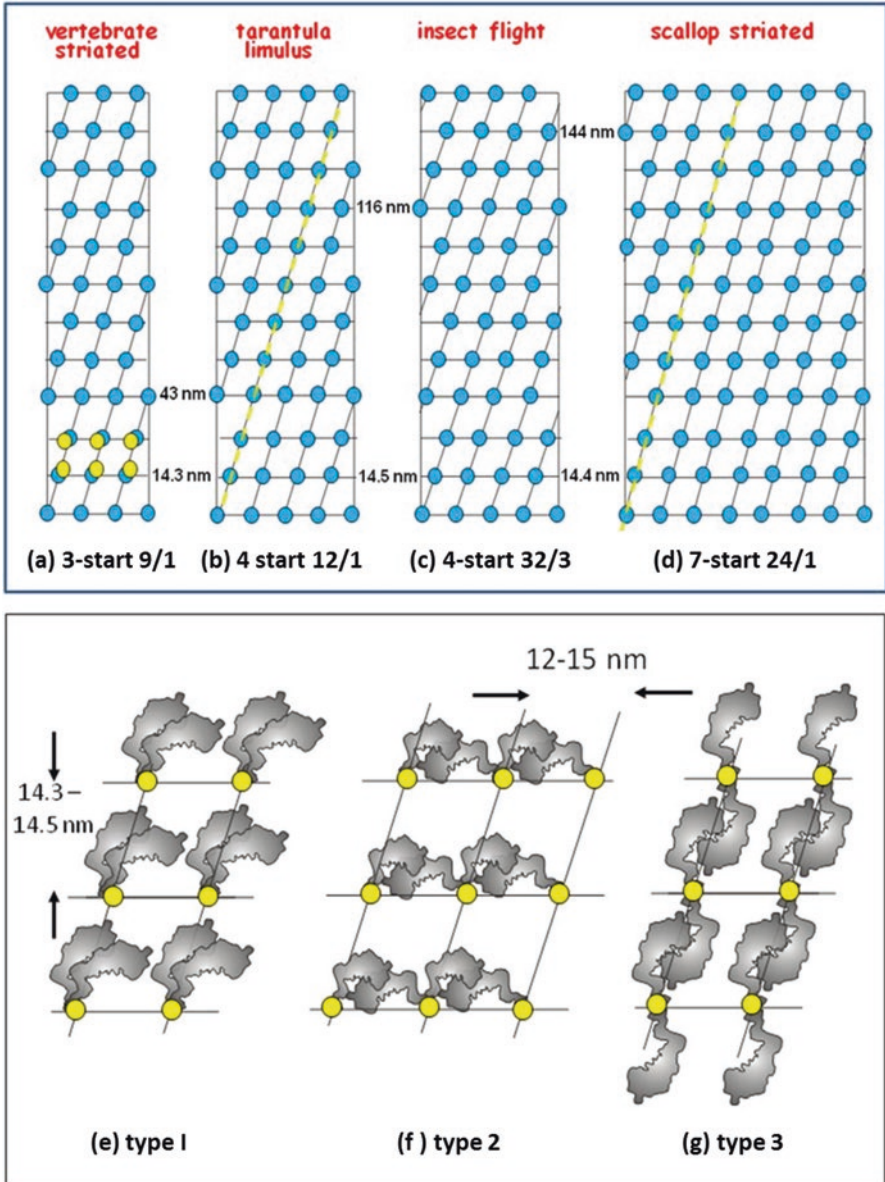
Fig. 11.14 (**a–d**) The known myosin head surface lattices of a variety of myosin filaments as labelled. The 14.3–14.5 nm crown repeat is a characteristic of all myosin filaments, but the angle that the long period strands follow (e.g. the *yellow dashed lines* in (**b**)) varies slightly to give the filaments different axial repeats. All vertebrate striated muscle bipolar filaments are 1.57 μm long, have threefold symmetry and have an axial repeat after three crowns of 43 nm. 4-start filaments can have either a 43 nm axial repeat (e.g. (**b**) tarantula, limulus) or a 115 nm repeat (c: insect flight muscle). Scallop thick filaments are the only known 7-start filaments. Their axial repeat is 144 nm. Note that the crowns in vertebrate striated muscles are not equivalent; there are small axial and
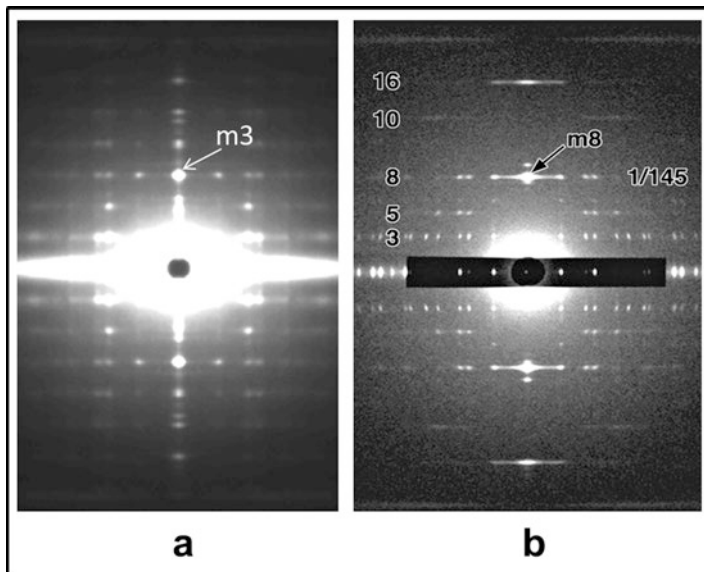
Fig. 11.15 Low-angle X-ray diffraction patterns from striated muscles held vertical. (**a**) The pattern from bony fish muscle and (**b**) the pattern from insect flight muscle. The meridional reflection from the crown spacing is m3 in (**a**), the 3rd order of the 43 nm repeat, and m8 in (**b**), the 8th order of the 115 nm repeat. These two muscle types are the only striated muscles known to have good 3D order as exemplified by the sampling of the (horizontal) layer lines by (vertical) row lines, giving well defined diffraction spots ((**a**) from the work of Harford and Squire 1986; (**b**) courtesy of Professor M.K. Reedy)

helical reconstruction methods (e.g. Stewart and Kensler 1986; Crowther et al. 1985; Offer et al. 2000 and many more). Filaments from most sources tended to show ridges of density along the long period helices in one of the nets in Fig. 11.14, and these were all interpreted as myosin heads in the type 3 conformation (Fig. 11.14g). Examples of some excellent reconstructions from scorpion and *Limulus* muscles (Stewart et al. 1985) are shown in Fig. 11.16, where it appeared that type 3 interactions might well occur. The problem with this and other similar studies was that the resolution involved was simply not quite good enough to be able to distinguish the alternative head configurations.

Fig. 11.14 (continued) azimuthal perturbations of the head arrays, as shown by the *yellow dots*. (**e** to **f**) Schematic illustrations of the kinds of head arrays that might occur in resting muscle myosin filaments if their organization is stabilised by head-to-head interactions. (**e**) Type 1 has heads in the same myosin molecule interacting. (**f**) Type 2 has heads from adjacent myosin molecules in a crown interacting. (**g**) Type 3 has heads from successive crowns interacting. For many years it was thought that the head arrangement in (**g**) occurred on many filament types, but this was based on 3D reconstructions where the resolution was not sufficient to distinguish between the alternatives. It is now known that type 1 interactions occur in all muscle myosin filaments so far studied in detail. (Adapted from Squire et al. 2005)
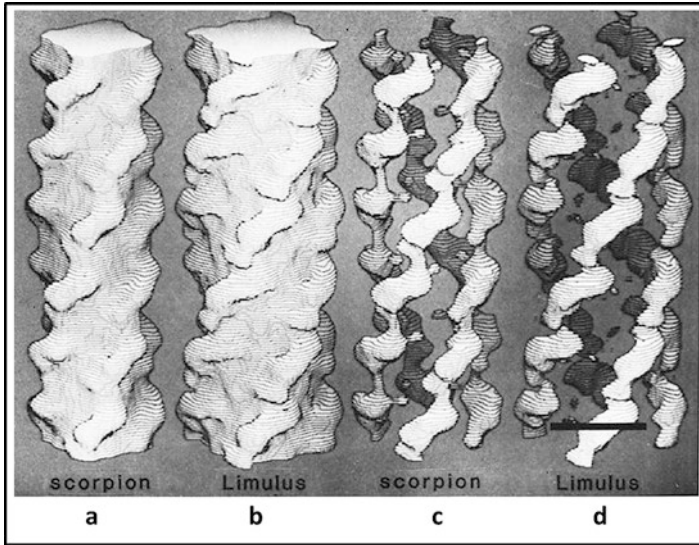
Fig. 11.16 Early 3D reconstructions of arthropod muscle thick filaments (**a**) and (**c**) scorpion; (**b**) and (**d**) *Limulus*. The reconstructions were obtained from negatively stained filaments and the helical reconstruction method (DeRosier and Moore 1971) at about 5 nm resolution. They show apparent axial spreading of head density along the long period helical tracks of the cross-bridge lattice (see Fig. 11.14b) (Reproduced from Stewart et al. 1985 with permission)

As in the case of actin filaments, the real breakthrough came when single particle methods were applied to these filaments (AL-Khayat et al. 2004). The first results from this were from the 4-start myosin filaments in tarantula muscle. These filaments have the advantage that the myosin heads appear to be organised on perfectly regular helices, so helical averaging could be applied. Tarantula filament images obtained from frozen-hydrated specimens were studied using single particle analysis and the Egelman (2000) helical averaging procedure, and the study of myosin filaments took a huge step forward (Fig. 11.17b; Woodhead et al. 2005). Here it was found that tarantula myosin heads, previously thought to be in the type 3 conformation, were actually in the type 1 conformation, with the two heads in one myosin molecule interacting with each other, but in slightly different conformations (Figs. 11.17a and 11.18). In a previous study of a 2D raft of myosin heads from vertebrate smooth muscle viewed in ice, Wendt et al. (2001) had seen a similar two-headed interaction with the motor domain of one head interacting with the converter domain of the other head. Smooth muscle myosin is regulated by phosphorylation of the regulatory light chain and this smooth muscle myosin conformation was that found in the absence of phosphorylation. These authors described a free head and a blocked head, but the conformation could be such that neither head can bind to actin. The fact that this same conformation with free and blocked heads (Fig. 11.18) appeared on the surface of tarantula myosin filaments gave the hint that this might be a general theme, at least for myosin-regulated muscles (Craig and Woodhead
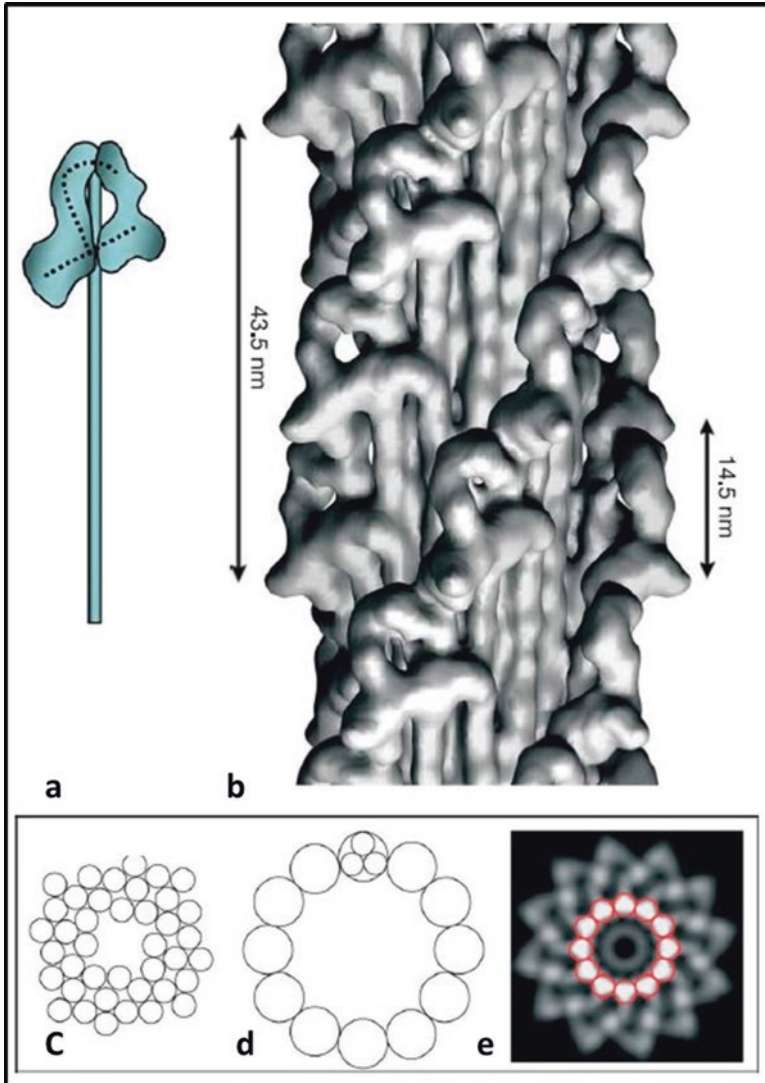
Fig. 11.17 (**b**) Surface view of a 3D reconstruction of tarantula muscle myosin filaments (bare zone at the *bottom*) by single particle analysis showing the interacting head motif illustrated schematically in (**a**) (Adapted and reoriented from Woodhead et al. (2005) to be consistent with other figures in the Chapter) (**c**–**e**) Possible models (**c**, **d**) and the average projection (**e**) of the cross-section ((**a**) and (**b**) Adapted from Woodhead et al. 2005; (**c**–**e**) from Craig and Woodhead 2006)

2006; Alamo et al. 2008). This idea was extended when the structures of the 4-start filaments in *Limulus* and the 7-start filaments in scallop striated muscle were studied by single particle analysis (Zhao et al. 2009; AL-Khayat et al. 2009a, b; Woodhead et al. 2013).

Fig. 11.18 Details of the
interacting head motif in
Fig. 11.17a, b with the
atomic model of vertebrate
smooth muscle heavy
meromyosin (PDB 1i84;
Wendt et al. 2001) fitted to
the density in the
reconstruction of tarantula
thick filaments (bare zone
to the bottom). Best fit of
the crystal structure shown
in (**b**) and the ribbon
diagram without the
density map shown in (**a**),
where particular features
of the heads are labelled
(*MD* motor domains, *ELC*
essential light chains, *RLC*
regulatory light chains)
(Modified and reoriented
from Woodhead et al.
(2005) to be consistent
with other figures in the
Chapter)
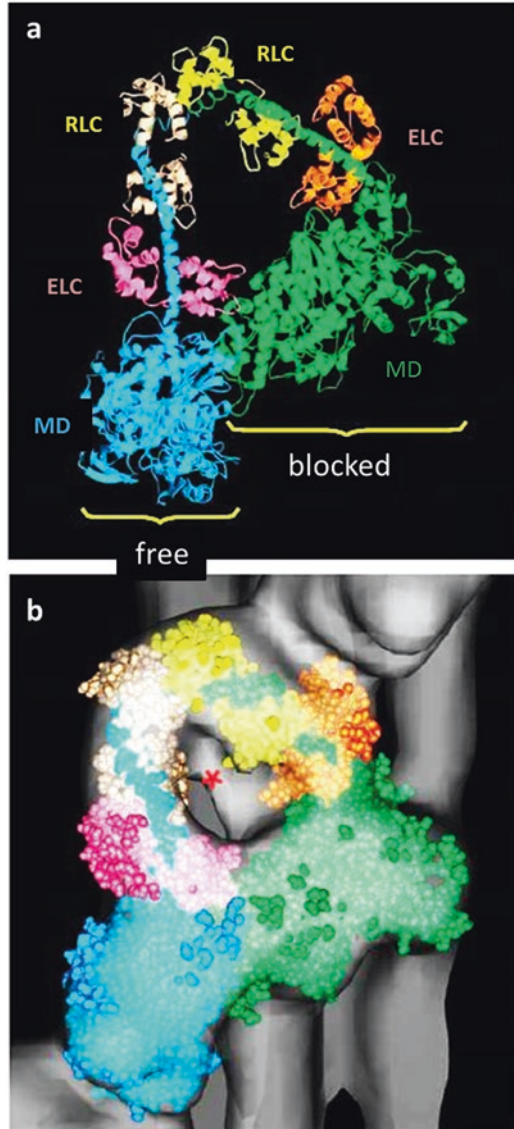
More surprising results came from similar analysis of vertebrate striated muscle
myosin filaments, which are thin filament-regulated (e.g. AL-Khayat et al. 2006).
What has not been described yet is the fact that the crowns in these vertebrate fila-
ments are not all the same; there are systematic perturbations of the myosin heads
away from strictly helical symmetry. Indications of axial and azimuthal head shifts
are shown in Fig. 11.14a. There is still approximately a true repeat after three crowns
or 42.9 nm, but the three crowns in this repeat are different. The axial perturbations
were first shown by Huxley and Brown (1967) for frog muscle and by Harford and

Squire (1986) for fish muscle looking at the meridian of the X-ray diffraction patterns, but they may also involve azimuthal displacements around the filament axis. This means that the helical-averaging single particle approach of Egelman (2000) is not applicable and, as in the case of the thin filament (Paul et al. 2004), to see the structure properly requires full single particle analysis without helical averaging (AL-Khayat et al. 2004).

11.3.3   Myosin Filaments in Vertebrate Striated Muscles: MyBP-C and Titin

Vertebrate striated muscle myosin filaments are bipolar with a bare zone half way along their length where there is antiparallel packing of myosin molecules (Figs. 11.1a and 11.13a, b). The myosin heads then appear in the two outer parts of the filaments, conveniently termed the bridge regions (shown schematically in Fig. 11.13b). However, even though there is a dominant 42.9 nm axial repeat, the bridge regions are not constant in structure, as a purely helical filament would be. As described in Chap. 10 (see also Granzier and Labeit, 2005, 2007), the giant molecule titin runs along the myosin filament from the M-band to the filament tip, where it forms so-called end-filaments (Trinick 1981), and then it continues through the I-band to the Z-line. The myosin filament part of titin, mostly composed of immunoglobulin (Ig) and fibronectin (Fn3)-like domains, has systematic pseudo-repeats along it as in Fig. 11.19a.

In the centre of each bridge region is a part of the filament where the extra protein C-protein (or myosin binding protein C; MyBP-C; Offer et al. 1973; Bennett et al. 1986) is located. This region, the C-zone (Sjostrom and Squire 1977), coincides with part of the titin sequence where there are repeats of sets of 11 domains. Successive domains are separated axially by about 3.9 nm on average, giving a repeat after 11 domains of 42.9 nm – exactly the same as the myosin axial repeat. MyBP-C in vertebrate skeletal muscle consists of 10 domains which, like much of titin, are Ig-like or Fn3-like (Fig. 11.19b). Originally, it was thought to be just a myosin-binding protein with domains C8 to C10 binding to myosin and to titin on the myosin filament. But more recently it has been found that the outer end of C-protein, around domains C1 and C2 in vertebrate skeletal muscle, binds to actin filaments (Squire et al. 2003; Whitten et al. 2008; Luther et al. 2011). The C-protein therefore acts as a bridge between the myosin and actin filaments, at least under some conditions (Fig. 11.19c; Luther and Craig 2011). Cardiac muscle MyBP-C has an extra domain (C0) at its N-terminus (Fig. 11.19b) and both skeletal and cardiac MyBP-C have a Pro-Ala-rich linker on the N-terminal side of C1, all of which may also be involved in actin binding (Squire et al. 2003). Another myosin-associated protein is H-protein (MyBP-H; Fig. 11.19b) which is also a molecule with Ig and Fn3-like domains; it is found closer to the bare zone than MyBP-C (Bennett et al. 1986). Other related proteins which cross-link myosin filaments in
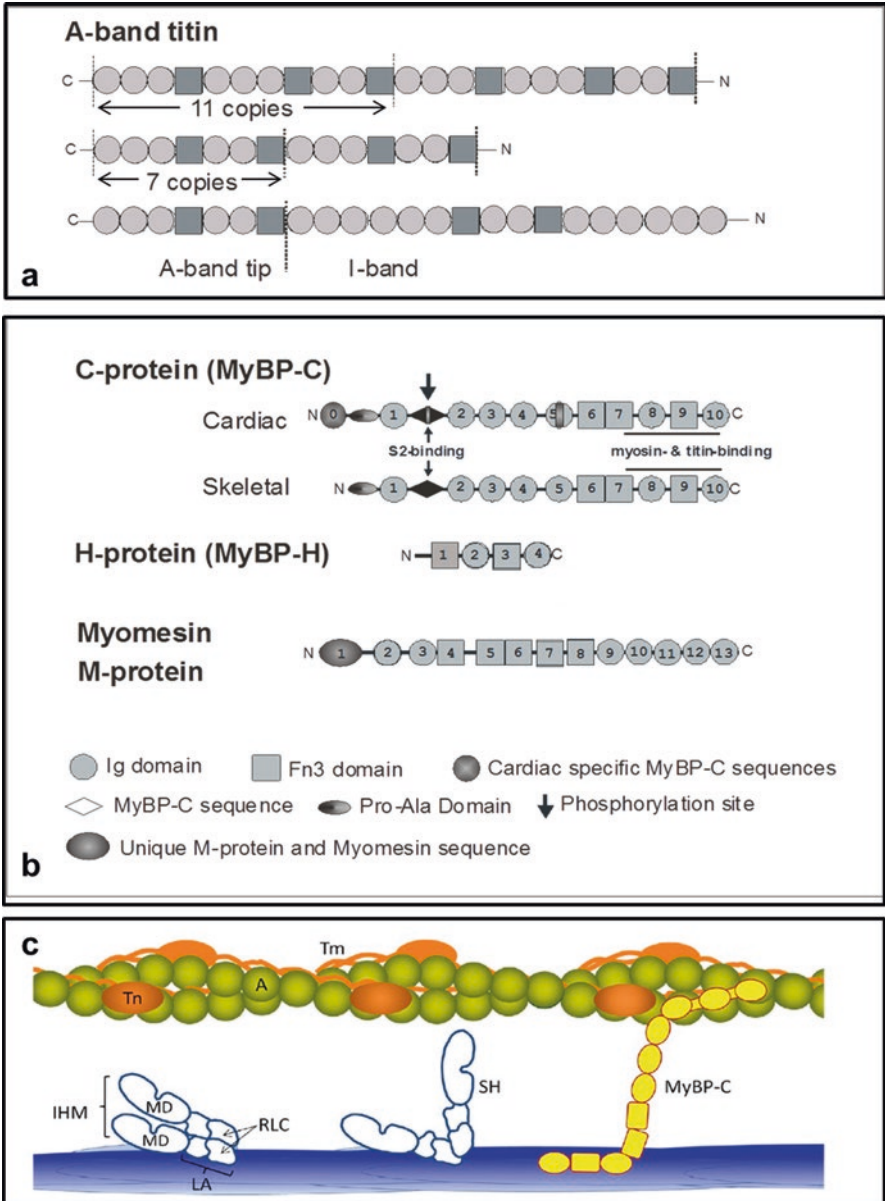
Fig. 11.19 (**a**) The A-band part of titin consists of a repeating pattern of 11 Ig and Fn3 domains grouped in 11 copies from C-terminus to N-terminus as (Fn-Fn-Fn-Ig-Fn-Fn-Fn-Ig-Fn-Fn-Ig) starting at the edge of the bare zone. It then continues as six copies of the repeat (Fn-Fn-Fn-Ig-Fn--Fn-Ig) plus a few extra domains out to the myosin filament tip. (**b**) C-protein (MyBP-C) in cardiac and skeletal muscle. The common features are ten domains, which are Fn3-like, and Ig-like, numbered C1 to C10. The C-terminal domains C8, C9, C10 are associated with myosin and titin binding. Between C1 and C2 is a sequence which binds to myosin S2. Cardiac C-protein has an

the M-band are myomesin and M-protein (Agarkova and Perriard 2005; Tskhovrebova and Trinick 2012; Fig. 11.19b).

For some years there was uncertainty about whether MyBP-C might have a different axial repeat from myosin, because various features of the meridional X-ray diffraction pattern (Rome et al. 1973) and analysis of longitudinal sections of vertebrate muscle A-bands (Squire et al. 1982) suggested that MyBP-C might have an axial repeat around 44 nm rather than 42.9 nm. It was then found that the longer repeat, which certainly did come from MyBP-C, was not the myosin repeat, but was due to the axial disposition of the actin ends of the MyBP-C molecules, while the myosin ends of MyBP-C were still on a 42.9 nm repeat (Squire et al. 2003). Interestingly, if the length of the C-zone was much more than seven to nine 42.9 nm repeats, then this apparent longer spacing for the actin-binding ends would have disappeared and reverted to 42.9 nm. The fact that MyBP-C can bind to actin under certain conditions leads one to wonder whether it has some sort of regulatory role in the contractile cycle, and this has led to much further work and discussion (Kunst et al. 2000; Kulikovskaya et al. 2003; Winegrad 2003; Luther et al. 2008; Ababou et al. 2008; Kensler and Harris 2008; Govada et al. 2008; Shaffer et al. 2009; Luther et al. 2011; Craig et al. 2014; Walcott et al. 2015; van Dijk et al. 2015; Previs et al. 2015; Moss et al. 2015; Kampourakis et al. 2015 and many others).

11.3.4 Myosin Filaments in Vertebrate Striated Muscles: The Myosin Head Array

The presence of MyBP-C in the C-zone and the fact that titin has a C-zone-specific sequence might suggest that the myosin head configurations in the C-zone are different from those in the other parts of the bridge region; the P-zone (P for proximal; between the bare zone and the C-zone) and the D-zone (D for distal; outside the C-zone) (Sjostrom and Squire 1977). Indeed, recent evidence from probe studies suggests that this is so (Fusi et al. 2015). A great deal of analysis of the low-angle X-ray diffraction data from the myosin filaments, particularly in fish muscles, where the A-band is highly ordered, was rigorous, but it was all based on the assumption that the myosin head organisation is the same along all of the vertebrate striated muscle bridge region (Fig. 11.15a); Harford and Squire 1986; Hudson et al. 1997; AL-Khayat and Squire 2006). On this assumption, the number of X-ray

---

Fig. 11.19 (continued) additional Ig domain (C0) at its N-terminus. Phosphorylation sites are indicated by a rectangle in the S2 binding site and in C5. A pro-Ala-rich domain is present at the N-terminus in both isoforms; it links C0 and C1 in the cardiac molecule. (**c**) Illustration of the cross-linking of the myosin and actin filaments by C-protein (MtBP-C) and possible interaction of 'sentinel' myosin heads (SH) with actin in relaxed muscle. *A* actin, *Tn* troponin, *Tm* tropomyosin, *MD* myosin motor domain, *RLC* regulatory light chain, *LA* lever arm, *IHM* interacting heads motif (cf Fig. 11.15), MyBP-C myosin binding protein C (C-protein) (part (c) reproducedfrom Woodhead and Craig 2015)

observations on the myosin layer lines out to a resolution of around 6 nm was greater than the number of parameters to fit in a model and unique solutions could be obtained. But X-ray diffraction provides information about average structures, so if the myosin head organisation varies along the vertebrate myosin filament then this poses a real problem for X-ray diffraction analysis. To model the P-zone, C-zone and D-zone heads as different structures would require more parameters than there are X-ray observations. Only if there was good information from another source would the analysis of vertebrate X-ray diffraction patterns be fully tractable.

As mentioned above, the way around this problem and around the fact that the filaments are not helical is to use single particle analysis without helical averaging. This has been done successfully for rabbit cardiac myosin filaments in negative stain (Zoghbi et al. 2008; AL-Khayat et al. 2008). An improved structure was then obtained by AL-Khayat et al. (2013) using just the C-zones in myosin filaments from the ideal source, the human heart, once again in negative stain. A particle was taken to be approximately two 42.9 nm repeats, with one of the MyBP-C stripes as the centre point (Fig. 11.20b). We know that these filaments have threefold rotational symmetry (Fig. 11.14a), so this could be used to help get a starting model from one of the class averages. An independent, reference-free, angular assignment method was also used (Padwardhan et al. 2004). The result is shown in Fig. 11.21a.

What is striking from this analysis and that of Zoghbi et al. (2008) is that the head pairs are in very similar arrangements to the head pairs on the tarantula and scallop thick filaments discussed earlier. Even though in vertebrate filaments the three crowns of heads are different (Fig. 11.21b), they are all in similar conformations; it is the angle of the whole head pair assembly that changes in the different crowns. Note that the views of the head pairs in Fig. 11.21b are from the inside of the filament where the end of the myosin rod can be seen inserting into the heads, whereas those in Fig. 11.18 are from the outside. There are still free heads and blocked heads, as in smooth muscle myosin (Wendt et al. 2001), but this time the myosin filaments are not directly myosin-regulated. The vertebrate striated muscle essential light chain can be phosphorylated reversibly during contraction, but this seems to be a modulating process rather than the fundamental regulatory switch. So the head pair arrangement appears to be a common theme in most types of myosin filament.

Figure 11.21a showing the human cardiac myosin filament also reveals densities which were interpreted as being due to titin and to MyBP-C. Each half myosin filament is thought to be associated with six titin strands (Liversage et al. 2001; Knupp et al. 2002) and it is likely that these occur in three pairs. The density in the 3D reconstruction appears to be able to accommodate such titin pairs and the 3.9 nm axial repeat is also evident, tending to confirm the density as being due to titin. At the top end of the Figure there is also enough extra density to account for domains C8 to C10 of MyBP-C, also with a 3.9 nm axial repeat, so we have a general picture of the head, titin and MyBP-C organisation in these cardiac thick filaments.
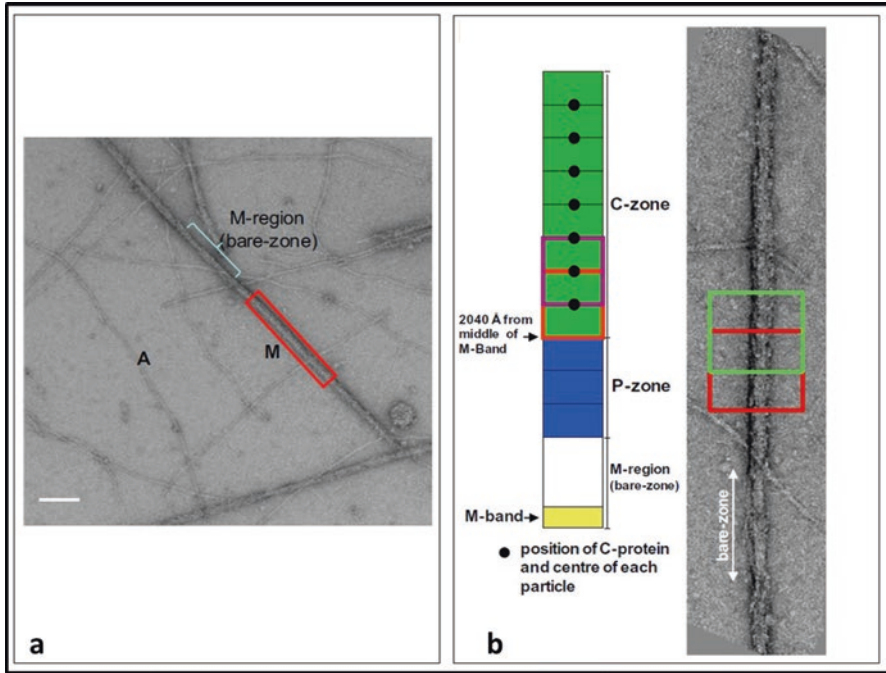
Fig. 11.20 (**a**) Electron micrograph of isolated myosin filaments (M) from the ventricular muscle of normal (undiseased) human heart in the relaxed state. Some actin filaments (A) can be seen in the background. Scale bar 200 nm. (**b**) left schematic diagram showing the different inner A-band regions within a half filament from the M-band at the bottom, the M-region, the P-zone and the C-zone (Sjostrom and Squire 1977). For the 3D reconstruction in Fig. 11.21 particles were selected from the C-zone only and the particles were about two C-repeats long (about 89 nm; *rectangular boxes* in (**b**)), but centred on each C-repeat (Reproduced from Al-Khayat et al. (2013) Supporting information with permission)

11.3.5 Myosin Filaments in Insect Flight Muscle

As in many aspects of its structure and contractile behaviour, when it comes to the myosin filaments in insect flight muscle they are in a class of their own. After the brilliant pioneering work of Reedy (1968) and that reported in his subsequent papers, it was found that the IFM myosin filaments have their heads on 4-start helices with a crown separation of 14.4 nm and a filament repeat after 115 nm (Fig. 11.14c). An early helical reconstruction of isolated IFM filaments showed this 4-strandedness and gave hints as to the head disposition (Morris et al. 1991). It was mentioned earlier that modelling low-angle X-ray diffraction data from vertebrate myosin filaments would be problematical if it turns out that the head organisation in the C-zone is not the same as that in the P-zone and D-zone. However, it appears that the IFM thick filaments are probably purely helical along most of their length. In this case, application of rigorous data stripping (Rajkumar et al. 2007) and

Fig. 11.21 (**a**) Map of a 3D reconstruction of human heart myosin filaments as in Fig. 11.20 showing a full 43 nm repeat consisting of three crowns of heads and fitted atomic structures for the myosin heads, titin (*yellow*) and C-protein (MyBP-C C8 to C10, *magenta*). The bare zone is towards the bottom. The myosin head pair organisations in the three crowns are almost identical and the same as that in Fig. 11.18 for tarantula (Woodhead et al. 2005) and vertebrate smooth muscle heavy meromyosin (Wendt et al. 2001). (**b**) The difference between crowns is that the similar head pair assemblies are rotated slightly differently (Reproduced from AL-Khayat et al. 2013 with permission)
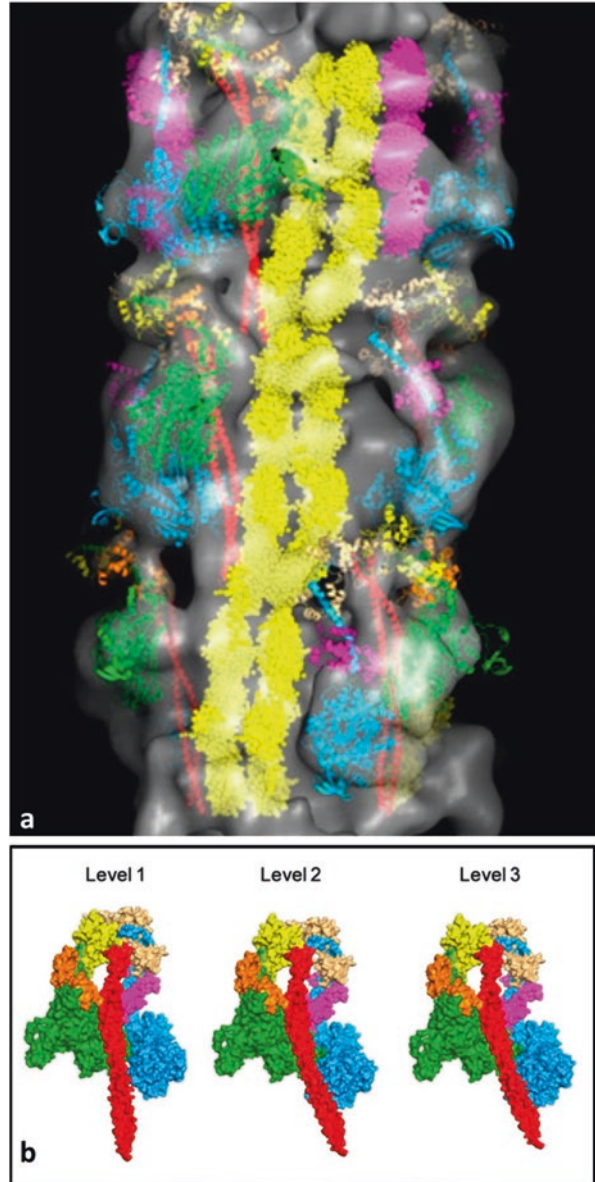
modelling methods (Hudson et al. 1997) to the beautiful IFM X-ray diffraction data of Reedy (Figs. 11.15b and 11.22a); AL-Khayat et al. 2003), assuming perfect helical symmetry, generated a plausible arrangement for the resting IFM myosin heads, which is quite different from the head pairs in all the other filaments that have been studied. The head pairs do not tilt towards the surface of the filament backbone as in
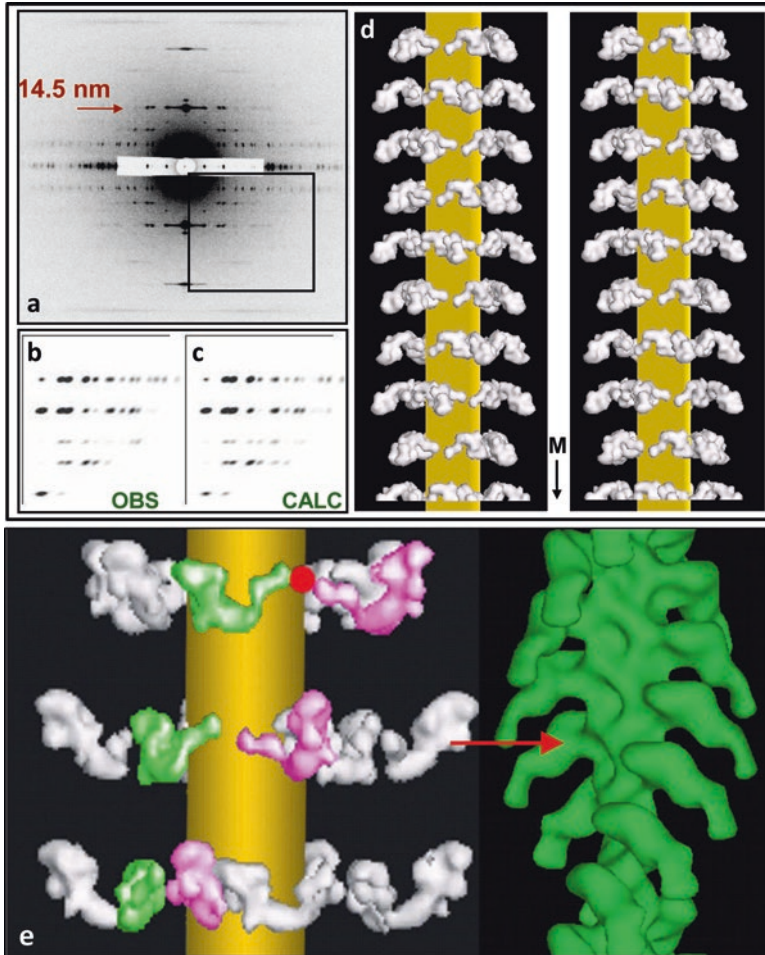
Fig. 11.22 (**a**) Low-angle X-ray diffraction pattern from insect flight muscle (*Lethocerus*) showing the lower right hand quadrant of peaks that are stripped as in (**b**) using CCP13 software ('*Fibrefix*'; Rajkumar et al. 2007) and modelled as in (**c**, **d**) by the method of Hudson et al. (1997). The calculated diffraction data (**c**) from the model in (**d**) can be compared to the original observed intensities in (**b**). (**e**) Illustration of the rather small change in motor domain orientation required for a resting head in insect flight muscle to interact with actin in rigor (*green* reconstruction to the right; Harford JJ and Squire JM unpublished data from S1-labelled fish muscle). In the insect muscle lattice the actin and myosin filaments would be much closer together, so only a small radial head movement is needed for it to attach to actin. The transition from an initial attached state to rigor would then require a large swing of the lever arm on a relatively static motor domain. Note that this figure is inverted relative to (**d**); the M-band here would be at the top (All figures reproduced from AL-Khayat et al. 2003 with permission)

other muscles, but they tilt out into a plane perpendicular to the filament axis. This kind of structure is supported by the fact that electron micrographs of IFM thick filaments show very marked ridges or shelves of density every 14.5 nm along the filaments, rather than density aligned along the long pitch helices as seen in other filament types. This is also consistent with the very strong 14.5 nm meridional reflection in the IFM diffraction patterns (Figs. 11.15b and 11.22a). The head pairs were tentatively thought to be in the type 2 configuration (Fig. 11.22), but this may be incorrect (see late breaking results, Sect. 11.5).

Unlike other muscles, asynchronous insect flight muscles, such as those in *Lethocerus* discussed here, are designed to oscillate when activated. They display the phenomenon of stretch activation whereby the muscles pull back harder when they are stretched so that together with the thorax/wing assembly to which they are connected they act as a resonant mechanical system; only the occasional nervous input is needed to keep the system oscillating. As shown in AL-Khayat et al. (2003), the resting myosin heads in insect flight muscle (Fig. 11.22e) appear to be in a configuration whereby they could attach to actin with only a small radial movement; the system appears suited to the rapid synchronous head movements necessary for oscillatory contraction.

11.3.6 The Myosin Filament Backbone

With enormous advances in knowledge of the myosin filament symmetries in different muscles and the different arrangements of the myosin heads, the remaining major unknown is how the coiled-coil myosin rods are packed into the filament backbone. We know from a variety of X-ray diffraction observations on the packing of two-chain coiled-coils that they are about 2 nm in diameter when hydrated and their centre-to-centre packing distance is about the same (e.g. Squire 1975; Elliott 1979). The original 'General Model of Myosin Filament Structure' by Squire (1971) was based on the premise that all myosin filament surface lattices would be rather similar, as they have proved to be (see Fig. 11.14), and this itself was based on the idea that myosin rods would presumably pack in a certain way to give rise to the related surface lattices. On this basis Squire (1973) proposed a packing scheme for all types of myosin filaments based on plausible rod interactions, equivalent packing and the ability to generate the known head surface lattices. He came up with a curved crystal model where the rods are fairly straight and are only slightly tilted from the long axis of the myosin filaments. An example of this arrangement for 4-stranded filaments is illustrated in Fig. 11.17c, where each circle represents the cross-section of one myosin rod, all myosin rods are equivalent and the array produces a perfectly helical surface lattice. The related structure for the 3-start filaments in vertebrate striated muscle is shown in more detail in Figs. 11.23d and 11.24a taken from Chew and Squire (1995).

This curved crystal model is not the only packing scheme that can make the myosin rods equivalent. There are also a variety of myosin sub-filament models in
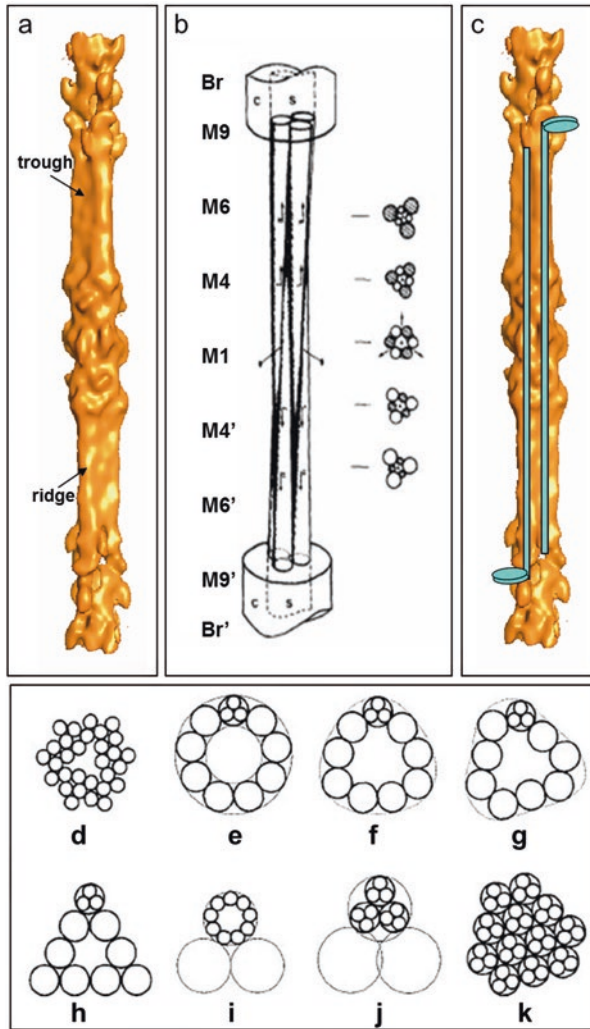
Fig. 11.23 Analysis of the bare zone of vertebrate striated muscle thick filaments. (**a**) shows a single particle reconstruction of the bare zone and start of the two bridge regions in fish skeletal muscle obtained by AL-Khayat et al. (2010). (**b**) Shows results based on an early analysis of the bare zone structure by Luther et al. (1981), suggesting that the backbone consists of three tapering densities from one end of the filament interleaved with three similar densities of opposite polarity coming from the other end of the filament. Various levels of extra M-band proteins are shown as M1 to M6, M6′ and the starts of the bridge regions are shown as Br and Br′. Results in (**a**) and (**b**) are entirely consistent. (**c**) shows how a myosin molecule (azure) might fit into the observed bare zone density. ((**a**–**c**) Reproduced from AL-Khayat et al. (2010) with permission). (**d**–**k**) Various schemes for the packing of myosin rods in the bridge regions of 3-start myosin filaments. (**d**) is a curved molecular crystal according to Squire (1973). (**e**–**k**) Various subfilament models for 3-start filaments. If the subfilaments have three myosin rods on average and all myosin molecules are equivalent then a large hollow ring is produced (**e**), which is not observed. The hollowness can be reduced by distorting the ring in various ways as in (**f**) to (**h**), but then the equivalence of the myosin rods is lost. The subfilaments could be larger as in (**i**), or the smaller subfilaments could be supercoiled as in (**j**) or close packed as in (**k**). Myosin rods in (**i**) and (**j**) would still be equivalent. Those in (**k**) would be far from equivalent ((**d**–**f**) Reproduced from Chew and Squire (1995) with permission)
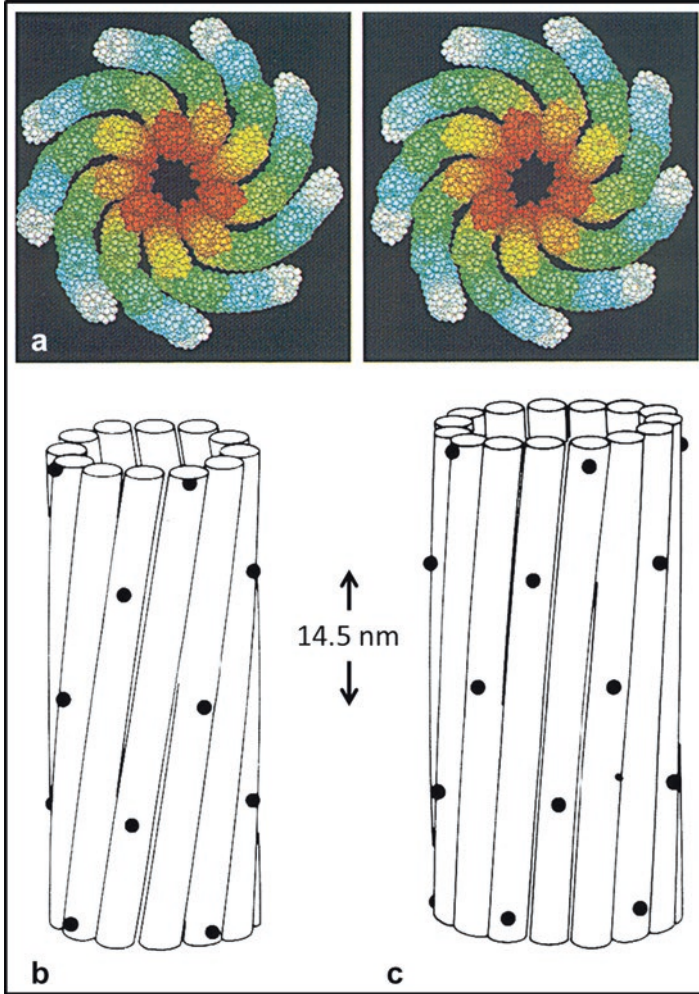
Fig. 11.24 (**a**) Stereo visualisation of the molecular crystal packing scheme proposed by Squire (1973) and shown as a simple diagram in Fig. 11.23d. Each amino acid is represented by a coloured sphere. The rods spiral down the filament at a very small angle. (**b**) and (**c**) Two possible structures for the myosin filaments in crustacean muscles (**a**) fast abdominal flexor of the lobster, (**b**) slow muscle of the lobster crusher claw. (**a**) is a 4-start filament as in other invertebrate muscles (Fig. 11.14b, c) and (**b**) a 5-start filament whose existence has not been confirmed. Sub-filament backbone structures are shown (Reproduced from Wray (1979) with permission)

which the myosin rods first pack into sub-filaments, and then a number of similar sub-filaments pack together to form the filament backbone (Squire 1975, 1981, 1986; Wray 1979). Examples of this kind of structure are shown in Figs. 11.23e–k and 11.24b, c. This is the structure preferred by Woodhead et al. (2005) on the basis of their 3D reconstruction of Tarantula myosin filaments (Fig. 11.17d, e). A charac-

teristic of these sub-filament models is that they generate large empty filament cores (they are hollow shells) unless there is an intermediate level of coiling as in Fig. 11.23j for a 3-stranded filament.

The problem with these different models is that the resolution of the best myosin filament reconstructions (Woodhead et al. 2005; AL-Khayat et al. 2013) is still only about 2.5 nm. Since the rods pack only 2 nm apart it is very difficult to determine the rod positions in the current reconstructions; real insight into myosin rod packing will only come from reconstructions with significantly higher resolution. Having said that, Chew and Squire (1995) suggested, after modelling high-angle (i.e. high resolution; 0.51 nm) X-ray diffraction data from vertebrate muscles, that the rods were very likely to be almost parallel to the muscle long axis. This would be true for the curved crystal model, but the rod tilt gets much bigger in some sub-filament models (Chew and Squire 1995).

11.3.7 *The Vertebrate Myosin Filament Bare Zone*

An intriguing aspect of bipolar myosin filaments is what happens in the bare zone, where the myosin rod packing is antiparallel. Squire (1973) made a suggestion about how the rods might pack to give the triangular filament profile that was observed in this region of the filament in cross-section. The story was then advanced substantially by Luther et al. (1981) in their rigorous analysis of frog muscle cross-sections, where they showed that the whole myosin filament has the symmetry of the dihedral point group 32. This means that, as well as the threefold rotational symmetry intrinsic to the 3-start helical arrangement of myosin heads (Fig. 11.14a), there are also three twofold rotation axes perpendicular to the long axis of the filament and exactly halfway along the filament in the middle of the bare zone. Rotation of the whole myosin filament about one of these axes by 180° would leave the filament appearance unchanged. This region was then visualised in 3D by AL-Khayat et al. (2010), the dihedral 32 point group symmetry was confirmed, and then this symmetry was applied to the reconstruction to provide further structural averaging. The final structure, Fig. 11.23a, b, was interpreted as showing three tapering densities more or less parallel to the filament long axis coming down from the top of the Figure, interdigitating with three similar but antiparallel densities coming up from the bottom end of the filament. Figure 11.23c shows how the rods might fit into these densities. This scheme would explain why filament cross-sections appear fairly triangular each side of the M-band (around lines M4 and M6 up to M9; Fig. 11.23b), but more circular in the middle at M1, where the six tapering densities, both up and down, all have the same diameter (Luther et al. 1981). Once again, more resolution is needed to see how the myosin rods are actually packed, but the implication from this structure is that the rods are fairly parallel to the filament long axis.

11.3.8 Paramyosin Filaments

The muscles of many molluscs are smooth muscles in the sense that their myosin and actin filaments are not organised into regular sarcomeres. In addition, many of these muscles have very large myosin-containing filaments. They are very much larger in diameter than in vertebrate or insect muscles and they can be very long. The bulk of these filaments is composed of the two-chain coiled-coil α–helical protein paramyosin, which is like the rod part of myosin, but with no myosin heads on the end (Kantha et al. 1990). High-angle X-ray diffraction patterns from these filaments have suggested that the 2 nm diameter paramyosin molecules pack into a nearly tetragonal body-centred lattice about 2.7 nm on a side. There was some debate about how these rods are packed on a larger scale. Elliott G.F (1964a) originally suggested on the basis of appearances in muscle sections that these filaments might have the molecules packed into flat layers. Elliott A. and Lowy (1970) then suggested that the structure night have a flat layer of molecules wrapped into a kind of swiss-roll structure – giving the filaments a helicoidal arrangement of paramyosin molecules. Further work by Elliott (A) and Bennett (1984) then confirmed that Elliott G.F. had been right and that there is a flat layer structure – a structure similar to synthetic aggregates studied by Cohen et al. (1971).

An early observation by X-ray diffraction (Bear and Selby 1956) later confirmed by electron microscopy (e.g. Szent-Gyorgyi et al. 1971) was that paramyosin filaments show a characteristic net-like surface arrangement (Fig. 11.25a, b), known as the Bear-Selby net. It has the characteristic axial repeat of 14.5 nm between crowns as in other myosin filaments, but here the surfaces are relatively flat and the Bear-Selby net is almost a 2D lattice. The repeating units – clusters of paramyosin molecules – are shifted axially by 14.5 nm and laterally by differing amounts on different paramyosin filaments to give a true axial repeat (the 'c' repeat; Fig. 11.24b) after five of the 14.5 nm crown levels, giving a repeat of c = 72 nm. The variable lateral spacing in the Bear-Selby net is usually termed the 'a' spacing (Fig. 11.25b). The results of Elliott (1979) and Elliott and Bennett (1984) clearly showed that if paramyosin filaments showing the Bear-Selby net pattern are rotated around their long axes then the net pattern tends to disappear and a simple 14.5 nm axial banding is obtained. Their preferred structure is as in Fig. 11.25c–e. Figure 11.25d shows how stain might accumulate between the molecular ends to give the stain patches which show up the Bear-Selby net (Fig. 11.25a).

Squire (1971) discussed what a layer of myosin molecules might be like on the surface of a roughly cylindrical backbone of paramyosin (Fig. 11.25e). Figure 11.26 shows the correlation between the lengths and diameters of paramyosin filaments and the molar ratio of paramyosin to myosin heavy chain as determined by Levine et al. (1976). The measurements appear to fit with a myosin layer of between 4 and 5 nm thickness as expected. The odd thing is that if the net is like one of the arrangements in Fig. 11.14a–d, but with a 72 nm axial repeat, the interaction which such a layer would make with the underlying paramyosin molecules would sometimes be
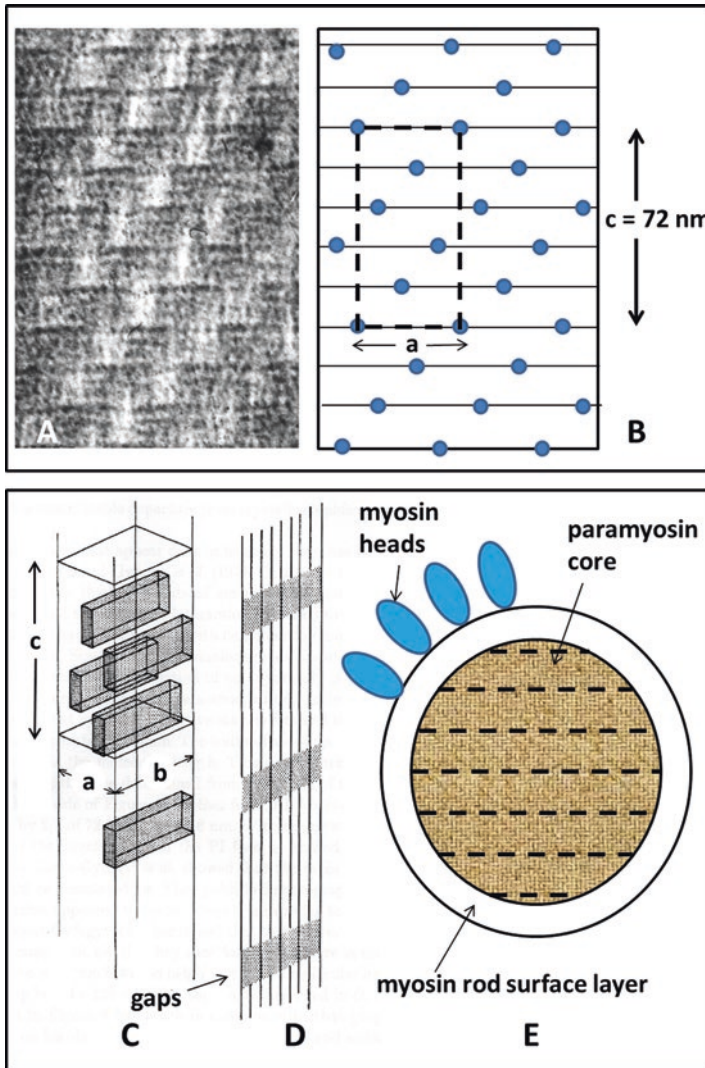
Fig. 11.25 (**a**) The surface lattice of paramyosin filaments as visualised in negative stain showing the Bear-Selby net pattern (Bear and Selby 1956). Features of the pattern are summarised in (**b**). (**c, d**) Three-dimensional perspective drawing of the paramyosin packing model of Elliott (1979) showing regions of stain extending through a filament. The 'b' direction is normal to the plane through 'a' and 'c'. (**b**) Diagrammatic perspective representation showing a single layer of paramyosin molecules from (**c**) each 129.1 nm long with an axial displacement of c = 72 nm between neighbours (as in Cohen et al. (1971)). Stain can lodge in gaps 72 nm apart between the ends of the molecules [(**c, d**) Reproduced from Elliott (1979) with permission]. (**e**) Illustration of a paramyosin filament cross-section with a central layered paramyosin structure and a surface layer of myosin rods from which myosin heads (*blue*) project
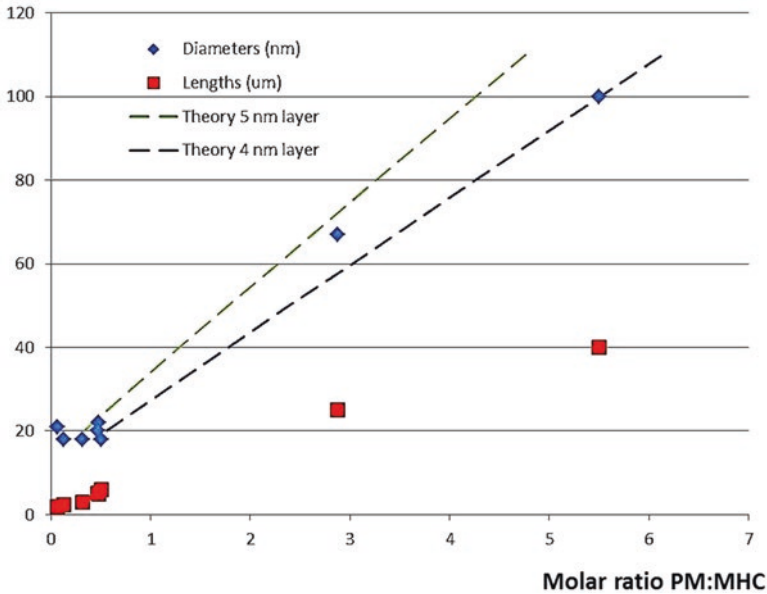
Fig. 11.26 Plots of the measured diameters and lengths of paramyosin filaments from a variety of sources plotted against the molar paramyosin to myosin heavy chain ratio as determined by Levine et al. (1976). The observed diameters (*blue diamonds*) are consistent with a myosin layer of thickness about 4–5 nm – *dashed lines*

with a Bear-Selby net structure and sometimes with a structure which has simple 14.5 nm axial banding. Is that plausible, or is it, for example, only the Bear-Selby net face that carries myosin?

11.3.9   *Myosin Filaments in Vertebrate Smooth Muscles*

The last two sections here are about myosin filaments that do not fit to the standard patterns found in other muscles as discussed so far. The first are the myosin filaments in vertebrate smooth muscles. In the 1950s and 1960s, when the first electron micrographs of striated muscles were obtained, vertebrate intestinal and other visceral muscles were also studied. It was confirmed that, unlike striated muscles, there were no sarcomeres present, but the surprising result was also found that there were no myosin filaments apparent in these muscles (e.g. Shoenberg 1969). Eventually it was assumed that vertebrate smooth muscle myosin filaments must be present when the muscles are active, but that they may assemble as needed from a pool of myosin molecules in inactive muscles. Attempts were then made to find conditions under which smooth muscles myosin molecules would assemble into filaments that could be visualised. The breakthrough came with the study of guinea pig *Taenia coli* muscles (from along the large intestine) when treated with hypertonic Ringer solution
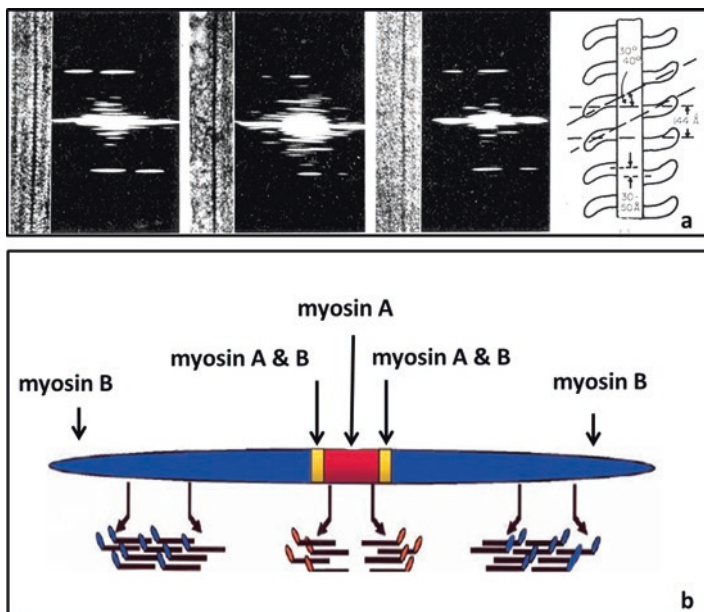
Fig. 11.27 (**a**) Myosin filaments (*left images* in each pair) in vertebrate smooth muscle seen in edge view and the corresponding asymmetrical optical diffraction pattern. This shows the presence of face polar (side polar) arrangements of projecting myosin heads in which heads on opposite faces point towards opposite ends of the filaments, as summarised in the *right hand figure* (Figure Reproduced from Small and Squire (1972) with permission). (**b**) The distribution of myosins A and B in the bipolar thick filaments of C. elegans body wall muscle (Modified from Landsverk and Epstein 2005)

(Lowy et al. 1970; Vibert et al. 1971). Under these conditions (slightly stronger than normal Ringer solution) a very clear 14.4 nm meridional X-reflection characteristic of the presence of myosin filaments was observed (Vibert et al. 1971). A meridional reflection at 0.51 nm, characteristic of oriented α–helical rods (presumably mainly myosin rods) was also seen. Then, in a major electron microscopy study of thin sections of *Taenia coli* (Small and Squire 1972), the myosin filaments themselves were observed and analysed. However, they were definitely not bipolar filaments as found in striated muscles. The myosin aggregates that were observed were ribbon-shaped and very long. They were called myosin ribbons. Image analysis of edge views of the sectioned filaments clearly showed that the projecting myosin heads on the 14.4 nm repeat were not pointing in the same direction on each side of the filament as would occur in a bipolar filament (Fig. 11.27a), but heads on one side of the filament pointed towards the opposite end of the ribbon to the heads on the other side (Fig 11.27a right hand image). This was termed face polarity and, as detailed in Small and Squire (1972), this structure enabled smooth muscles to operate over a very large range of muscle lengths, since the interacting actin filaments, which were very long, would move in opposite directions on opposite faces of the myosin ribbons. The face polarity was confirmed by image analysis of the edge view ribbon

projections where a characteristic optical diffraction pattern was obtained. Instead of showing peaks on each side of the meridional reflection at 14.4 nm, as might be expected for a bipolar filament, the ribbons gave a pattern where the 14.4 nm peak had an additional reflection on the same layer line but on one side of the meridian only. Looking at the top part of the pattern, if the extra peak on the 14.4 nm layer line was on the left, then on the bottom it was on the right (Fig. 11.27a).

Since that time it has been found that the filaments in this muscle are normally not as wide as the ribbons seen by Small and Squire (e.g. Somlyo et al. 1973), but the same kind of opposite polarity on opposite faces of smaller filaments was seen in aggregates of pure smooth muscle myosin (Craig and Megerman 1977; Xu et al. 1996). These smaller filaments with a roughly square backbone profile were termed side-polar, but this was an unnecessary and misleading name. A square prism has faces, so face-polarity would remain the correct term irrespective of the filament width. The physiological advantages of this kind of polarity, even if the filaments are termed side-polar, are identical to those detailed by Small and Squire (1972). The smaller filaments may just aggregate under some conditions to produce the wider ribbons (e.g. Somlyo et al. 1973).

Small and Squire (1972) originally thought it possible that the ribbons might have a backbone core consisting of the 10 nm (intermediate) filaments that were seen in abundance in the *Taenia coli* sections. However, this is not the case since filaments with side- or face- polarity can be generated from a pure preparation of smooth muscle myosin (Craig and Megerman 1977). The presence of face-polarity rather than bipolarity can also be explained in terms of the different amino acid sequences in myosins from striated and smooth muscles (Straussman et al. 2005). The intermediate filaments form a separate cytoskeletal network in smooth muscle cells (Small and Sobieszek 1977).

A recent survey of face-polar or side-polar filaments in vertebrate smooth muscles has shown considerable variability in length between muscles and muscle states, and the suggestion was made that filaments are in equilibrium with myosin dimers and that they assemble or disassemble as required (Liu et al. 2013).

11.3.10   *Nematode and* Limulus *Muscles*

The nematode *Caenorhabditis elegans* (C. elegans) has great advantages in carrying out genetic studies and has been widely used for a variety of purposes (Brenner 1974). Along with *Drosophila* and Zebra fish, it provides a method of genetic dissection and in vivo development in the study of muscle structure and behaviour (e.g. Barral and Epstein 1999; Landsverk and Epstein 2005). The thick filaments in the body wall muscle have two types of myosin with heavy chains A and B, together with paramyosin. The two myosins are distributed in a particular way in the bipolar thick filament (Fig. 11.27b). The central region is entirely myosin A, there are then

small transition regions with both myosins A and B, and then the peripheral parts of the filament are entirely composed of myosin B. The structure of the filament backbone and the head organisation are not really known, although some suggestions have been made (e.g. Müller et al. 2001). Further work is needed to position this filament into the family of other thick filaments.

The story of *Limulus* telson muscle thick filaments is rather different. The thick filaments appear relatively normal compared to those in other muscles in that they are bipolar and their heads lie on 4-start helices (Figs. 11.14b and 11.16b; Wray et al. 1974; Stewart et al. 1981; Zhao et al. 2009). What is different about these filaments is how they behave as the sarcomere length in the muscles changes. An advantage for some muscles is that they operate over a wide range of sarcomere lengths. With the normal ideas about the muscle sarcomere and the sliding filament mechanism (Fig. 11.1a), only a limited range of filament sliding can occur. This is especially true on shortening of the muscle when first the actin filaments penetrate the M-band and then, on further shortening, the myosin filaments collide with the Z-band. However, there is evidence that the myosin filaments in *Limulus* are designed in such a way that the myosin filament tips collapse by losing molecules when they approach the Z-band, thus allowing the *Limulus* muscles to shorten more than usual and therefore to operate over a wider sarcomere length range than normal (Dewey et al. 1977). No doubt further anomalous behaviours will become apparent when other myosin filaments are studied.

11.4 Future Prospects

Enormous progress has been made since Huxley first isolated myosin filaments from muscle. We now know the lattice symmetries and head arrangements in a number of muscles in some detail, but we still do not know how the myosin rods pack (but see late breaking news below). In the case of the thin filament, results are starting to emerge on the arrangement of troponin and the individual tropomyosin repeats within a regulatory unit. But, as in many such studies, the big thing is to improve the resolution of the reconstructions that are produced. Only then will we know how these filaments are made and how they work.

Recent developments in detector technology already allow for the determination of the structure of protein complexes by cryo-electron microscopy at a resolution as high as 2.2 Å (Bartesaghi et al. 2015), and it is likely that even higher resolution structures of appropriate complexes will be obtained in the near future. It seems probable, therefore, that cryo-electron microscopy and single particle analysis could be used to obtain much higher resolution structures than are currently available for both thick and thin muscle filaments, although this will also require substantial improvements in sample preparation.

11.5   Late Breaking Results

After the bulk of this chapter was written, the authors heard of the ground-breaking work of Hu et al. (2016) who have used electron microscopy and single particle analysis to define the structure of the thick filaments from insect flight muscle (*Lethocerus*). They found that the myosin head arrangement at about 2 nm resolution was exactly the same as in other muscles in being of the type shown in Figs. 11.18 and 11.21b. But as in Fig. 11.22, the head pairs were not close to the filament surface, but were angled so that the plane of the head density was fairly flat on each 14.5 nm spaced crown to give the shelves of density characteristic of insect flight muscle thick filaments. Even more remarkable in this work is that the backbone structure is revealed at about 0.6 nm resolution. This is sufficient to see the individual myosin rods. They found that at intermediate resolution there appeared to be sub-filaments, as has been suggested before, However, at the full 0.6 nm resolution, there were no sub-filaments and the rod packing was exactly as in Fig. 11.17c for a 4-stranded filament (Squire 1973). It is likely that other kinds of myosin filament also have this kind of molecular crystal packing.

References

Ababou A, Rostkova E, Mistry S, Le Masurier C, Gautel M, Pfuhl M (2008) Myosin binding protein C positioned to play a key role in regulation of muscle contraction: structure and interactions of domain C1. J Mol Biol 384:615–630

Agarkova I, Perriard JC (2005) The M-band: an elastic web that crosslinks thick filaments in the center of the sarcomere. Trends Cell Biol 15:477–485

Alamo L, Wriggers W, Pinto A, Bartoli F, Salazar L, Zhao FQ, Craig R, Padron R (2008) Three-dimensional reconstruction of tarantula myosin filaments suggests how phosphorylation may regulate myosin activity. J Mol Biol 384:780–797

AL-Khayat HA, Squire JM (2006) Refined structure of bony fish muscle myosin filaments from low-angle X-ray diffraction data. J Struct Biol 155:218–229

AL-Khayat HA, Yagi N, Squire JM (1995) Structural changes in actin-tropomyosin during muscle regulation: computer modelling of low-angle X-ray diffraction data. J Mol Biol 252:611–632

AL-Khayat HA, Hudson L, Reedy MK, Irving TC, Squire JM (2003) Myosin head configuration in relaxed insect flight muscle: X-ray modelled resting cross-bridges in a pre-powerstroke state are poised for actin binding. Biophys J 85:1063–1079

AL-Khayat HA, Morris EP, Squire JM (2004) Single particle analysis: a new approach to solving the 3D structure of myosin filaments. J Mus Res Cell Motil 25:635–644

AL-Khayat HA, Morris EP, Powell AS, Kensler RW, Squire JM (2006) 3D structure of vertebrate (fish) muscle myosin filaments by single particle analysis. J Struct Biol **155**:202–217

AL-Khayat HA, Morris EP, Kensler RW, Squire JM (2008) 3D structure of relaxed mammalian (rabbit) cardiac muscle myosin filaments by electron microscopy and single particle analysis. J Struct Biol 163:117–126

AL-Khayat HA, Morris EP, Squire JM (2009a) The 7-stranded structure of relaxed scallop muscle myosin filaments: support for a common head configuration in myosin-regulated muscles. J Struct Biol **166**:183–194

AL-Khayat HA, Morris EP, Squire JM (2009b) 3D structure of relaxed scallop striated muscle myosin filaments by electron microscopy and single particle analysis. J Struct Biol 166:183–194

AL-Khayat HA, Kensler RW, Morris EP, Squire JM (2010) Three-dimensional structure of the M-region (bare zone) of vertebrate striated muscle myosin filaments by single-particle analysis. J Mol Biol 403:763–776

AL-Khayat HA, Kensler RW, Squire JM, Marston SB, Morris EP (2013) Atomic model of the human cardiac muscle myosin filament. Proc Natl Acad Sci U S A 110:318–323

Barral JM, Epstein HF (1999) Protein machines and self-assembly in muscle organization. BioEssays 21:813–823

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science 348:1147–1151

Bear RS, Selby CC (1956) The structure of paramyosin fibrils according to X-ray diffraction. J Biophys Biochem Cytol 2:55–69

Behrmann E, Muller M, Penczek PA, Mennherz HG, Manstein DJ, Raunser S (2012) Structure of the rigor actin-tropomyosin-myosin complex. Cell 150:327–338

Bennett P, Craig R, Starr R, Offer G (1986) The ultrastructural location of C-protein, X-protein and H-protein in rabbit muscle. J Muscle Res Cell Motil 7:550–567

Bremel RD, Weber A (1972) Cooperation within actin filament in vertebrate skeletal muscle. Nat New Biol 238:97–101

Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77:71–94

Brown JH, Cohen C (2005) Regulation of muscle contraction by tropomyosin and troponin: how structure illuminates function. 'Fibrous Proteins: Muscle & Molecular Motors' (Ed Squire JM, Parry DAD). Adv Protein Chem 71:121–159.

Caspar DL, Cohen C, Longley W (1969) Tropomyosin: crystal structure, polymorphism and molecular interactions. J Mol Biol 41:87–107

Chew MW, Squire JM (1995) Packing of alpha-helical coiled-coil myosin rods in vertebrate muscle thick filaments. J Struct Biol 115:233–249

Clarke M, Hoffman W, Wray JS (1986) ATP binding and crossbridge structure in muscle. J Mol Biol 191:581–585

Cohen C, Szent-Györgyi AG, Kendrick-Jones J (1971) Paramyosin and the filaments of molluscan "catch" muscles. I. Paramyosin: structure and assembly. J Mol Biol 56:223–227

Craig R, Lehman W (2001) Crossbridge and tropomyosin positions observed in native, interacting thick and thin filaments. J Mol Biol 311:1027–1036

Craig R, Megerman J (1977) Assembly of smooth muscle myosin into side polar filaments. J Cell Biol 75:990–996

Craig R, Woodhead J (2006) Structure and function of myosin filaments. Curr Opin Struct Biol 16:204–212

Craig R, Lee KH, Mun JY, Torre I, Luther PK (2014) Structure, sarcomeric organization, and thin filament binding of cardiac myosin-binding protein-C. Pflugers Arch 466:425–431

Craig R, Padron R, Alamo L (1991) Direct determination of myosin filament symmetry in scallop striated adductor muscle by rapid freezing and freeze substitution. J Mol Biol 220:125–132

Crowther RA, Padron R, Craig R (1985) Arrangement of the heads of myosin in relaxed thick filaments from tarantula muscle. J Mol Biol 182:429–439

DeRosier DJ, Klug A (1968) Reconstruction of three-dimensional structures from electron micrographs. Nature 217:130–134

DeRosier DJ, Moore P (1971) Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. J Mol Biol 52:355–369

Dewey MM, Walcott B, Colflesh DE, Terry H, Levine RJC (1977) Changes in thick filament length in Limulus striated muscle. J Cell Biol 75:366–380

Dominguez R, Freyzon Y, Trybus KM, Cohen C (1998) Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-powerstroke state. Cell 94:559–571

Ebashi S, Endo M (1968) Calcium ion and muscle contraction. Prog Biophys Mol Biol 18:123–183

Egelman EH (2000) A robust algorithm for the reconstruction of helical filaments using single-particle methods. Ultramicroscopy 85:225–234

Elliott GF (1964a) Electron microscope studies of the structure of the filaments in the opaque adductor muscle of the oyster *Crassostrea Angulata*. J Mol Biol 10:89–104

Elliott GF (1964b) X-ray diffraction studies on striated and smooth muscles. Proc R Soc Lond B Biol Sci 160:467–472

Elliott GF (1967) Variations of the contractile apparatus in smooth and striated muscles: X-ray diffraction studies at rest and in contraction. J Gen Physiol 50(Suppl):171–184

Elliott A (1979) Structure of molluscan thick filaments: a common origin for diverse appearances. J Mol Biol 132:323–340

Elliott A, Bennett P (1984) Molecular organization of paramyosin in the core of molluscan thick filaments. J Mol Biol 176:477–493

Elliott A, Lowy J (1970) A model for the coarse structure of paramyosin filaments. J Mol Biol 53:181–203

Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116:190–199

Fujii T, Iwane AH, Yanagida T, Namba K (2010) Direct visualization of secondary structures of F-actin by electron cryomicroscopy. Nature 467:724–728

Fusi L, Huang Z, Irving M (2015) The conformation of myosin heads in relaxed skeletal muscle: implications for myosin-based regulation. Biophys J 109:783–792

Govada L, Carpenter L, da Fonseca PCA, Helliwell JR, Rizkallah P, Flashman E, Chayen NE, Redwood C, Squire JM (2008) Crystal structure of the C1 domain of cardiac myosin binding protein-C: implications for hypertrophic cardiomyopathy. J Mol Biol 378:387–397

Granzier HL, Labeit S (2005) Titin and its associated proteins. In 'Fibrous Proteins: Muscle & Molecular Motors' (Ed Squire JM, Parry DAD). Adv Protein Chem 71: 89–119.

Granzier HL, Labeit S (2007) Structure-function relations of the giant elastic protein titin in striated and smooth muscle cells. Muscle Nerve 36:740–755

Hanson J, Lowy J (1963) The structure of F-actin and of actin filaments isolated from muscle. J Mol Biol 6:46–60

Hardwicke PM, Hanson J (1971) Separation of thick and thin myofilaments. J Mol Biol 59:509–516

Harford J, Squire JM (1986) "Crystalline" myosin cross-bridge array in relaxed bony fish muscle. Low-angle X-ray diffraction from plaice fin muscle and its interpretation. Biophys J 50:145–155

Haselgrove JC (1972) X-ray evidence for a conformational change in actin-containing filaments of vertebrate striated muscle. Cold Spring Harb Symp Quant Biol 37:341–352

Herzberg O, James MN (1988) Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 A resolution. J Mol Biol 203:761–779

Houdusse A, Szent-Gyorgyi AG, Cohen C (2000) Three conformational states of scallop myosin S1. Proc Natl Acad Sci 97:11238–11243

Holmes KC, Lehman W (2008) Gestalt-binding of tropomyosin to actin filaments. J Muscle Res Cell Motil 29:213–219

Holmes KC, Popp D, Gebhard W, Kabsch W (1990) Atomic model of the actin filament. Nature 347:44–49

Holmes KC, Schröder RR, Sweeney HL, Houdusse A (2004) The structure of the rigor complex and its implications for the power stroke. Philos Trans R Soc Lond Ser B Biol Sci 359:1819–1828

Hu Z, Taylor DW, Reedy MK, Edwards RJ, Taylor KA (2016) Structure of myosin filaments from relaxed Lethocerus flight muscle by cryo-em at 6 Å resolution. Sci Adv 2(9):e1600058

Hudson L, Harford JJ, Denny RC, Squire JM (1997) Myosin head configuration in relaxed fish muscle: resting state myosin heads must swing axially by up to 150 Å or turn upside down to reach rigor. J Mol Biol 273:440–455

Huxley AF, Niedergerke R (1953) Structural changes in muscle during contraction: interference microscopy of living muscle fibres. Nature 173:971–973

Huxley HE (1963) Electron microscope studies on the structure of natural and synthetic protein filaments from striated muscle. J Mol Biol 7:281–308

Huxley HE (1969) The mechanism of muscular contraction. Science 164:1356–1365

Huxley HE (1972) Structural changes in actin- and myosin-containing filaments during contraction. Cold Spring Harb Symp Quant Biol 37:361–376

Huxley HE, Hanson EJ (1953) Structural basis of the cross-striations in muscle. Nature 172:530–532

Huxley HE, Brown W (1967) The low-angle X-ray diagram of vertebrate striated muscle and its behaviour during contraction and rigor. J Mol Biol 30:383–434

Jung SK, Komatsu S, Ikebe M, Craig R (2008b) Head-head and head-tail interaction: a general mechanism for switching off myosin II activity in cells. Mol Biol Cell 19:3234–3242

Jung SK, Burgess SA, Billington N, Colegrave M, Patel H, Chalovich JM, Chantler PD, Knight PJ (2008a) Conservation of the regulated structure of folded myosin 2 in species separated by at least 600 million years of independent evolution. PNAS 105:6022–6026

Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC (1990) Atomic structure of the actin: DNase I complex. Nature 347:37–44

Kampourakis T, Yan Z, Gautel M, Sun YB, Irving M (2015) Myosin binding protein-C activates thin filaments and inhibits thick filaments in heart muscle cells. Proc Natl Acad Sci U S A 111:18763–18768

Kantha SS, Watabe S, Hashimoto K (1990) Comparative biochemistry of paramyosin – a review. J Food Biochem 14:61–88

Kensler RW, Harris SP (2008) The structure of isolated cardiac myosin thick filaments from cardiac myosin binding protein C knockout mice. Biophys J 94:1707–1718

Kensler RW, Stewart M (1983) Frog skeletal muscle thick filaments are three-stranded. J Cell Biol 96:1797–1802

Knupp C, Luther PK, Squire JM (2002) Titin organisation and the 3D architecture of the vertebrate-striated muscle I-band. J Mol Biol 322:731–739

Kuhlbrandt W (2014) The resolution revolution. Science 343:1443–1444

Kulikovskaya I, McClellan G, Flavigny J, Carrier L, Winegrad S (2003) Effect of MyBP-C binding to actin on contractility in heart muscle. J Gen Physiol 122:761–774

Kunst G, Kress KR, Gruen M, Uttenweiler D, Gautel M, Fink RH (2000) Myosin binding protein C, a phosphorylation-dependent force regulator in muscle that controls the attachment of myosin heads by its interaction with myosin S2. Circ Res 86:51–58

Labeit S, Ottenheijm CA, Granzier H (2011) Nebulin, a major player in muscle health and disease. FASEB J 25:822–829

Landsverk ML, Epstein HF (2005) Genetic analysis of myosin II assembly and organization in model organisms. Cell Motil Life Sci 62:2270–2282

Lehman W, Craig R, Vibert P (1994) Ca(2+)-induced tropomyosin movement in Limulus thin filaments revealed by three-dimensional reconstruction. Nature 368:65–67

Lehman W, Vibert P, Uman P, Craig R (1995) Steric-blocking by tropomyosin visualized in relaxed vertebrate muscle thin filaments. J Mol Biol 251:191–196

Lehrer SS, Morris EP (1982) Dual effects of tropomyosin and troponin-tropomyosin on actomyosin subfragment 1 ATPase. J Biol Chem 257:8073–8080

Lehrer SS, Geeves MA (1998) The muscle thin filament as a classical cooperative/allosteric regulatory system. J Mol Biol 277:1081–1089

Levine RJC, Elfvin M, Dewey MM, Walcott B (1976) Paramyosin in invertebrate muscles II Content in relation to structure and function. J Cell Biol 71:273–279

Li XE, Tobacman LS, Mun JY, Craig R, Fischer S, Lehman W (2011) Tropomyosin position on F-actin revealed by EM reconstruction and computational chemistry. Biophys J 100:1005–1013

Liu JCY, Rottler J, Wang L, Zhang J, Pascoe CD, Lan B, Norris BA, Herrera AM, Paré PD, Seow CY (2013) Myosin filaments in smooth muscle cells do not have a constant length. J Physiol 591:5867–5878

Liversage AD, Holmes D, Knight PJ, Tskhovrebova L, Trinick J (2001) Titin and the sarcomere symmetry paradox. J Mol Biol 305:401–409

Lowey S, Slayter HS, Weeds AG, Baker H (1969) Substructure of the myosin molecule. I. Subfragments of myosin by enzymic degradation. J Mol Biol 42:1–29

Lowy J, Hanson J (1962) Ultrastructure of invertebrate smooth muscles. Physiol Rev Suppl 5:34–47

Lowy J, Poulsen FR, Vibert J (1970) Myosin filaments in vertebrate smooth muscle. Nature 225:1053–1054

Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semi-automated software for high-resolution single-particle reconstructions. J Struct Biol 128:82–97

Luther PK, Craig R (2011) Modulation of striated muscle contraction by binding of myosin binding protein C to actin. BioArchitecture 1:277–283

Luther PK, Munro PMG, Squire JM (1981) Three-dimensional structure of the vertebrate muscle A-band III: M-region structure and myosin filament symmetry. J Mol Biol 151:703–730

Luther PK, Bennett PM, Knupp C, Craig R, Padrón R, Harris SP, Patel J, Moss RL (2008) Understanding the organisation and role of myosin binding protein C in normal striated muscle by comparison with MyBP-C knockout cardiac muscle. J Mol Biol 384:60–72

Luther PK, Winkler H, Taylor K, Zoghbi ME, Craig R, Padrón R, Squire JM, Liu J (2011) Direct visualization of myosin-binding protein C bridging myosin and actin filaments in intact muscle. Proc Natl Acad Sci U S A 108:11423–11428

Lymn RW, Taylor EW (1971) Mechanism of adenosine triphosphate hydrolysis by actomyosin. Biochemistry 10:4617–4624

Manning EP, Tardiff JC, Schwartz SD (2011) A model of calcium activation of the cardiac thin filament. Biochemistry 50:7405–7413

Maw MC, Rowe AJ (1980) Fraying of A-filaments into three subfilaments. Nature 286:412–414

McLachlan AD, Stewart M (1976) The 14-fold periodicity in alpha-tropomyosin and the interaction with actin. J Mol Biol 103:271–298

Menetret JF, Schroder RR, Hofmann W (1990) Cryo-electron microscopic studies of relaxed striated muscle thick filaments. J Mus Res Cell Motil 11:1–11

McElhinny AS, Kolmerer B, Fowler VM, Labeit S, Gregorio CC (2001) The N-terminal end of nebulin interacts with tropomodulin at the pointed ends of the thin filaments. J Biol Chem 276:583–592

McKillop DF, Geeves MA (1993) Regulation of the interaction between actin and myosin subfragment 1: evidence for three states of the thin filament. Biophys J 65:693–701

Millman BM, Bennett PM (1976) Structure of the cross-striated adductor muscle of the scallop. J Mol Biol 103:439–467

Mok, NS. (2005) X-ray diffraction studies of defined states in fish muscle. PhD Thesis, Imperial College London.

Moore PB, Huxley HE, DeRosier DJ (1970) Three-dimensional reconstruction of F-actin, thin filaments and decorated thin filaments. J Mol Biol 50:279–295

Morris EP, Squire JM, Fuller GW (1991) The 4-stranded helical arrangement of myosin heads on insect (*Lethocerus*) flight muscle thick filaments. J Struct Biol 107:237–249

Moss RL, Fitzsimons DP, Ralphe JC (2015) Cardiac MyBP-C regulates the rate and force of contraction in mammalian myocardium. Circ Res 116:183–192

Müller SA, Häner M, Ortiz I, Aebi U, Epstein HF (2001) STEM analysis of Caenorhabditis elegans muscle thick filaments: evidence for microdifferentiated substructures. J Mol Biol 305:1035–1044

Offer G, Moos C, Starr R (1973) A new protein of the thick filaments of vertebrate skeletal myofi-brils. Extractions, purification and characterization. J Mol Biol 74:653–676

Offer G, Knight PJ, Burgess S, Alamo L, Padron R (2000) A new model for the surface arrange-ment of myosin molecules in tarantula thick filaments. J Mol Biol 298:239–260

Ohtsuki I (1979) Molecular arrangement of troponin-T in the thin filament. J Biochem 86:491–497

Ottenheijm CA, Buck D, de Winter JM, Ferrara C, Piroddi N, Tesi C, Jasper JR, Malik FI, Meng H, Stienen GJ, Beggs AH, Labeit S, Poggesi C, Lawlor MW, Granzier H (2013) Deleting exon 55 from the nebulin gene induces severe muscle weakness in a mouse model for nemaline myopathy. Brain 136:1718–1731

Otterbein LR, Graceffa P, Dominguez R (2001) The crystal structure of uncomplexed actin in the ADP state. Science 293:708–711

Parry DAD (1975) Analysis of the primary sequence of alpha-tropomyosin from rabbit skeletal muscle. J Mol Biol 98:519–535

Parry DAD, Squire JM (1973) Structural role of tropomyosin in muscle regulation: analysis of the X-ray diffraction patterns from relaxed and contracting muscles. J Mol Biol 75:33–55

Patwardhan A, Paul D, HA AL-K, EP M (2004) A measure for the angle between projections based on the extent of correlation between corresponding central sections. J Mol Biol 344:707–724

Paul DM, Patwardhan A, Squire JM, Morris EP (2004) Single particle analysis of filamentous and highly elongated macromolecular assemblies. J Struct Biol 148:236–250

Paul DM, Squire JM, Morris EP (2017) Relaxed and active thin filament structures reveal a new regulatory mechanism: varying tropomyosin shifts within a regulatory unit. J Struct Biol (in press)

Phillips GN Jr, Fillers JP, Cohen C (1986) Tropomyosin crystal structure and muscle regulation. J Mol Biol 192:111–131

Pirani A, Xu C, Hatch V, Craig R, Tobacman LS, Lehman W (2005) Single particle analysis of relaxed and activated muscle thin filaments. J Mol Biol 346:761–772

Poole KJ, Lorenz M, Evans G, Rosenbaum G, Pirani A, Craig R, Tobacman LS, Lehman W, Holmes KC (2006) A comparison of muscle thin filament models obtained from electron microscopy reconstructions and low-angle X-ray fibre diagrams from non-overlap muscle. J Struct Biol 155:273–284

Previs MJ, Prosser BL, Mun JY, Previs SB, Gulick J, Lee K, Robbins J, Craig R, Lederer WJ, Warshaw DM (2015) Myosin-binding protein C corrects an intrinsic inhomogeneity in cardiac excitation-contraction coupling. Sci Adv 1:e1400205

Rajkumar G, HA AL-K, Eakins F, Knupp C, JM S (2007) The CCP13 Fibrefix program suite: semi-automated analysis of diffraction patterns from non-crystalline materials. J Appl Crystallogr 40:178–184

Rayment I, Holden HM, Whittaker M, Yohn M, Lorenz M, KC H, RA M (1993a) Structure of the actin-myosin complex and its implications for muscle contraction. Science 261:58–65

Rayment I, Rypniewski WR, Schmidt-Base K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM (1993b) Three-dimensional structure of myosin subfragment-1: a molecular motor. Science 261:50–58

Reedy MK (1968) Ultrastructure of insect flight muscle. I. Screw sense and structural grouping in the rigor cross-bridge lattice. J Mol Biol 31:155–176

Reedy MK, Leonard KR, Freeman R, Arad T (1981) Thick myofilament mass determination by electron scattering measurements with scanning transmission electron microscopy. J Muscle Res Cell Motil 2:45–64

Rome E, Offer G, Pepe FA (1973) X-ray diffraction of muscle labelled with antibody to C-protein. Nat New Biol 244:152–154

Shaffer JF, Kensler RW, Harris SP (2009) The myosin binding protein C motif binds to F-actin in a phosphorylation sensitive manner. J Biol Chem 284:12318–12327

Shoenberg CF (1969) An electron microscope study of the influence of divalent ions on myosin filament formation in chicken gizzard extracts and homogenates. Tissue Cell 1:83–96

Sjostrom M, Squire JM (1977) Fine structure of the A-band in cryo-sections. The structure of the A-band of human skeletal muscle fibres from ultra-thin cryo-sections negatively stained. J Mol Biol 109:49–68

Slayter HS, Lowey S (1967) Substructure of the myosin molecule as visualized by electron microscopy. Proc Natl Acad Sci U S A 58:1611–1618

Small JV, Sobieszek A (1977) Studies on the function and composition of the 10-nm (100 Å) filaments of vertebrate smooth muscle. J Cell Sci 23:243–268

Small JV, Squire JM (1972) Structural basis of contraction in vertebrate smooth muscle. J Mol Biol 67:117–149

Somlyo AP, Devine CE, Somlyo AV, Rice RV (1973) Filament organization in vertebrate smooth muscle. Philos Trans R Soc Lond Ser B Biol Sci 265:223–229

Squire JM (1971) General model for the structure of all myosin-containing filaments. Nature 233:457–462

Squire JM (1972) General model of myosin filament structure. II. Myosin filaments and crossbridge interactions in vertebrate striated and insect flight muscles. J Mol Biol 72:125–138

Squire JM (1973) General model of myosin filament structure. III. Molecular packing arrangements in myosin filaments. J Mol Biol 77:291–323

Squire JM (1975) Muscle filament structure and muscle contraction. Annu Rev Biophys Bioeng 4:137–163

Squire JM (1981) The structural basis of muscular contraction. Plenum Press, New York

Squire JM (1986) Muscle myosin filaments: Internal structure and crossbridge organisation. Comments Mol Cell Biophys 3:155–177

Squire JM (2016) Muscle contraction: sliding filament history, sarcomere dynamics and the two Huxleys. Glob Cardiol Sci Pract 11:1–23 (http://dx.doi.org/10.21542/gcsp.2016.11)

Squire JM, Morris EP (1998) A new look at thin filament regulation in vertebrate skeletal muscle. FASEB J 12:761–771

Squire JM, Harford JJ, Edman AC, Sjostrom M (1982) Fine structure of the A-band in cryo-sections III: crossbridge distribution and the axial structure of the human C-zone. J Mol Biol 155:467–494

Squire JM, Luther PK, Knupp C (2003) Structural evidence for the interaction of C-protein (MyBP-C) with actin and sequence identification of a possible actin-binding domain. J Mol Biol 331:713–724

Squire JM, AL-Khayat HA, Knupp C, Luther PK (2005) 3D Molecular architecture of muscle. In 'Fibrous Proteins: Muscle & Molecular Motors'. Adv Protein Chem 71: 17–87.

Stewart M, Kensler RW (1986) Arrangement of myosin heads in relaxed thick filaments from frog skeletal muscle. J Mol Biol 192:831–851

Stewart M, Kensler RW, Levine RJ (1981) Structure of Limulus telson muscle thick filaments. J Mol Biol 153:781–790

Stewart M, Kensler RW, Levine RJ (1985) Three-dimensional reconstruction of thick filaments from Limulus and scorpion muscle. J Cell Biol 101:402–411

Straussman R, Squire JM, Ben-Yaakov A, Ravid S (2005) Molecular packing of myosin II rods relates directly to the linear sequence of charged amino acids. J Mol Biol 353:613–628

Szent-Györgyi AG, Cohen C, Kendrick-Jones J (1971) Paramyosin and the filaments of molluscan "catch" muscles. II. Native filaments: isolation and characterization. J Mol Biol 56:239–258

Takeda S, Yamashita A, Maeda K, Maeda Y (2003) Structure of the core domain of human cardiac troponin in the Ca(2+)-saturated form. Nature 424:35–41

Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A (2008) Protein structure fitting and refinement guided by cryo-EM density. Structure 6:295–307

Trinick JA (1981) End-filaments: a new structural element of vertebrate skeletal muscle thick filaments. J Mol Biol 151:309–314

Tskhovrebova L, Trinick JA (2003) Titin: properties and family relationships. Nat Rev Mol Cell Biol 4:679–689

Tskhovrebova L, Trinick JA (2012) Making muscle elastic: the structural basis of myomesin stretching. PLoS Biol 10(2):e1001264

van Dijk SJ, Witt CC, Harris SP (2015) Normal cardiac contraction in mice lacking the proline-alanine rich region and C1 domain of cardiac myosin binding protein C. J Mol Cell Cardiol 88:124–132

van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. J Struct Biol 116:17–24

van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A (2000) Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys 33:307–369

Vibert PJ (1992) Helical reconstruction of frozen-hydrated scallop myosin filaments. J Mol Biol 223:661–671

Vibert PJ, Craig R (1983) Electron microscopy and image analysis of filaments from scallop striated muscle. J Mol Biol 165:303–320

Vibert PJ, Haselgrove JC, Lowy J, Poulsen FR (1972) Structural changes in actin-containing filaments of muscle. J Mol Biol 236:182–183

Vibert P, Craig R, Lehman W (1997) Steric-model for activation of muscle thin filaments. J Mol Biol 266:8–14

Vinogradova MV, Stone DB, Malanina GG, Karatzaferi C, Cooke R, Mendelson RA, Fletterick RJ (2005) Ca(2+)-regulated structural changes in troponin. Proc Natl Acad Sci U S A 102:5038–5043

von der Ecken J, Muller M, Lehman W, Manstein DJ, Penczek PA, Raunser S (2015) Structure of the F-actin-tropomyosin complex. Nature 519:114–117

Walcott S, Docken S, Harris SP (2015) Effects of cardiac myosin binding protein-C on actin motility are explained with a drag-activation-competition model. Biophys J 108:10–13

Wendt T, Taylor D, Trybus KM, Taylor K (2001) Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. Proc Natl Acad Sci U S A 98:165–4366

Winegrad S (2003) Myosin binding protein C (MyBP-C) in cardiac muscle and contractility. Adv Exp Med Biol 538:31–40

Whitten AE, Jeffries CM, Harris SP, Trewhella J (2008) Cardiac myosin-binding protein C decorates F-actin: implications for cardiac function. Proc Natl Acad Sci U S A 105:18360–18365

Woodhead JL, Craig R (2015) Through thick and thin: interfilament communication in muscle. Biophys J 109:665–667

Woodhead JL, Zhao FQ, Craig R, Egelman EH, Alamo L, Padrón R (2005) Atomic model of a myosin filament in the relaxed state. Nature 436:1195–1199

Woodhead JL, Zhao FQ, Craig R (2013) Structural basis of the relaxed state of a Ca2+-regulated myosin filament and its evolutionary implications. Proc Natl Acad Sci U S A 110:8561–8566

Wray JS (1979) Structure of the backbone in myosin filaments of muscle. Nature 277:37–40

Wray JS, Vibert PJ, Cohen C (1974) Cross-bridge arrangements in Limulus muscle. J Mol Biol 88:343–348

Xu J-Q, Harder BA, Uman P, Craig R (1996) Myosin filament structure in vertebrate smooth muscle. J Cell Biol 134:53–66

Xu C, Craig R, Tobacman L, Horowitz R, Lehman W (1999) Tropomyosin positions in regulated thin filaments revealed by cryoelectron microscopy. Biophys J 77:985–992

Zhao F-Q, Craig R, Woodhead JL (2009) Head–head interaction characterizes the relaxed state of Limulus muscle myosin filaments. J Mol Biol 385:423–431

Zoghbi ME, Woodhead JL, Moss RL, Craig R (2008) Three-dimensional structure of vertebrate cardiac muscle myosin filaments. Proc Natl Acad Sci U S A 105:2386–2390

Chapter 12
Dystrophin and Spectrin, Two Highly Dissimilar Sisters of the Same Family

Olivier Delalande, Aleksander Czogalla, Jean-François Hubert, Aleksander Sikorski, and Elisabeth Le Rumeur

Contents

Abstract Dystrophin and Spectrin are two proteins essential for the organization of the cytoskeleton and for the stabilization of membrane cells. The comparison of these two sister proteins, and with the dystrophin homologue utrophin, enables us to emphasise that, despite a similar topology with common subdomains and a common structural basis of a three-helix coiled-coil, they show a large range of dissimilarities in terms of genetics, cell expression and higher level structural organisation. Interactions with cellular partners, including proteins and membrane phospholipids, also show both strikingly similar and very different behaviours. The differences between dystrophin and spectrin are also illustrated by the large variety of pathological anomalies emerging from the dysfunction or the absence of these proteins,

O. Delalande (✉) • J.-F. Hubert • E. Le Rumeur
Institut de Génétique et Développement de Rennes, UMR CNRS 6290, Université de Rennes 1, Rennes, France
e-mail: olivier.delalande@univ-rennes1.fr

A. Czogalla • A. Sikorski
Biotechnology Faculty, Department of Cytobiochemistry, University of Wrocław, ul. joliot-Curie 14a, 50-383 Wroclaw, Poland

373

showing that they are keystones in their function of providing a scaffold that sustains cell structure.

12.1   Introduction

Dystrophin is a unique protein coded by a single gene, which is expressed in differentiated smooth and striated muscular and nervous cells. Spectrins, on the other hand, are a large family of several proteins, coded by several genes, leading to a high number of isoforms that are ubiquitously expressed in numerous cell types. Dystrophin and its homologue utrophin are each coded by a single gene, *DMD* and *UTRN* (Table 12.1) respectively. These two genes are associated with the expression of unique proteins (Koenig et al. 1987; Tinsley et al. 1992). However, in the case of dystrophin, several promoters allow the tissue-specific expression of shorter isoforms of dystrophin compared to the full length protein. Utrophin is only known as the full length transcript (Helliwell et al. 1992; Tinsley et al. 1992). Unlike the two single genes for dystrophin and utrophin, spectrins are coded by numerous genes. In mammals, two genes, *SPTA1* and *SPTAN1,* code the two major α isoforms of spectrin, αI and αII, respectively. αI is characteristic of erythroid cells (Sahr et al. 1990). In contrast to *SPTA1*, alternative processing of the transcript of the *SPTAN1* gene gives rise to at least four isoforms of αII spectrin (Cianci et al. 1999). Moreover, four genes (*SPTB, SPTBN1, SPTBN2, SPTBN4*) code for conventional β spectrins (β-erythrocytic and βI, II and IV, respectively; Winkelmann et al. 1990; Stankewich et al. 1998; Berghs et al. 2000) and one gene (*SPTBN5*) codes for the large βV spectrin (Stabach and Morrow 2000). Variations in mRNA splicing in the case of isoforms βI, βII and βIV generate longer or shorter C-terminal regions (Berghs et al. 2000; Hayes et al. 2000; Table 12.1). The expression of the isoforms is tightly regulated in a tissue- and time-dependent manner, which is a consequence of the

Table 12.1 Genes coding dystrophin, utrophin and spectrins in *Homo sapiens*

Gene	Chromosome	Protein
DMD	X	Dystrophin
UTRN	6	Utrophin
SPTA1	1	Spectrin α erythrocytic (αI)
SPTAN1	9	Spectrin α non-erythrocytic (αII)
SPTB	14	Spectrin β erythrocytic
SPTBN1	2	Spectrin β non-erythrocytic (βI)
SPTBN2	11	Spectrin β non-erythrocytic 2 (βII)
SPTBN4	19	Spectrin β non-erythrocytic 4 (βIV)
SPTBN5	15	Spectrin β non-erythrocytic 5 (βV)

structural complexity of mammalian cells. Various intracellular locations (e.g. plasma membrane, Golgi apparatus, endoplasmic reticulum, nucleus) of spectrins reflect the variety of their functions. Only three spectrin isoforms are present in *D. melanogaster* and *C. elegans*, where the α subunit is encoded by a gene resembling mammalian *SPTAN1* and the β subunits correspond to mammalian βII and βV (Dubreuil and Grushko 1998).

Dystrophin, utrophin and the spectrins share the unique fold of the so-called "spectrin repeat", while they also have specific additional domains. Due to their similar spectrin repeats, they have been attributed to the same vast "spectrin" family of proteins (Parry et al. 1992). The spectrin family also includes α-actinin (see Ribiero Ede et al. 2014) which we will not discuss further in this Chapter. However, the general topology of dystrophin, utrophin and the spectrins is different, with dystrophin and utrophin existing as monomers and spectrins as homodimers or higher homo-oligomers of αβ-heterodimers. Their genes are all subject to mutations, but mutation types are very specific for each gene, with the vast majority being deletions of exons for dystrophin and point mutations for spectrins. The effect of out-of-frame mutations in the *DMD* gene is, in most cases, to cause the total absence of dystrophin in the musculature, which is highly deleterious and leads to Duchenne muscular dystrophy (DMD), with patients having an expected life span of less than 30 years. The effects of in-frame mutations are highly specific for the *DMD* gene and can lead to a shortened dystrophin, which partially maintains the dystrophin function and leads in most cases to Becker muscular dystrophy (BMD), with an expected life span that is very variable. In-frame point mutations of spectrins primarily affect the tetramerization processes.

Until now, the functions of dystrophin, utrophin and the spectrin proteins have not been well defined at the molecular level. They are largely thought to be scaffolding proteins interacting with numerous partners, some of them well documented, but there are probably many unknown partners. Their scaffolding function means that they are an important part of the structure of the cell cytoskeleton. However, understanding their function at a molecular level is only in its infancy, partly due to the large size of these proteins: from 71 to 427 kDa for the longest isoform of dystrophin and ranging between 246 and 430 kDa for spectrins.

12.2 Dystrophin, Utrophin and Spectrin Structural Domains

Dystrophin is a monomeric protein comprising 3685 residues in the longest isoform, which is present in muscle and heart cells (Fig. 12.1: Koenig et al. 1987, 1988). Several promotors allow for the tissue-specific expression of isoforms of various lengths. There are three full length isoforms. Dp427m is expressed in striated muscles, where it appears at the internal face of the sarcolemma. Dp427c is found in heart and glial cells (Koenig et al. 1989) and in brain and retina (Nudel et al. 1989; Chelly et al. 1990). Dp427p occurs in Purkinje cells and in skeletal
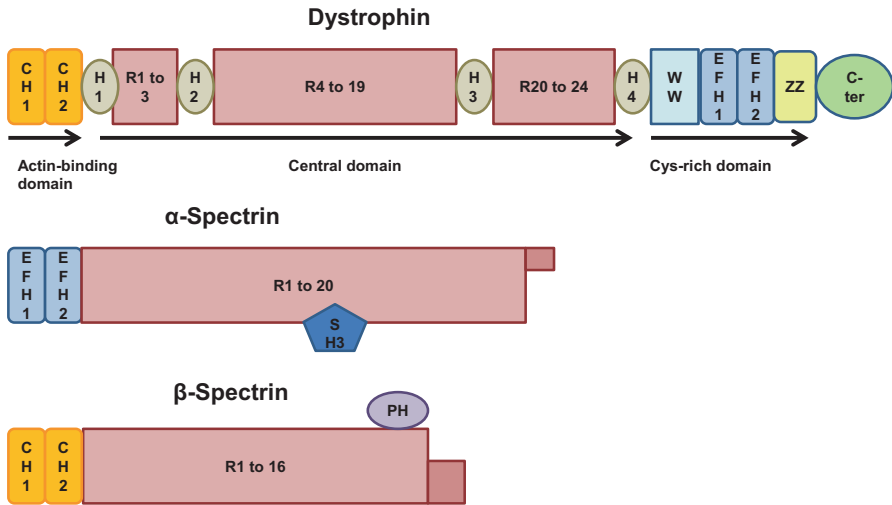
Fig. 12.1 Dystrophin and spectrin structural domains. CH1 and CH2, calponin homology domains, constituting the actin-binding domain ABD1; H1–H4, unstructured domains called hinges 1–4; R1–24, spectrin repeats, constituting the central domain; WW, Tryptophan-rich domain, EFH1 and EFH2, EF hand domains 1 and 2 and ZZ, ZZ domain, constituting the Cys-rich domain; C-ter, C-terminal domain, SH3, SH3 domain, PH, PH domain

muscle (Chelly et al. 1990). The shorter isoform, Dp160, is expressed in retina (D'Souza et al. 1995), Dp140 in brain, retina and kidney (Lidov et al. 1995), Dp116 in Schwann cells (Byers et al. 1993) and Dp71 in all tissues except skeletal muscles (Hugnot et al. 1992).

Dystrophin comprises four main domains (Le Rumeur et al. 2010). At the N-terminal end there is a first actin binding domain (ABD1) consisting of two calponin homology (CH) domains. It is followed by a very long central domain made of 24 spectrin repeats of about 100–110 residues each. They are interrupted by hinges, one between repeat 3 and 4 and one between repeats 19 and 20. A Cys-rich domain comprises WW, EFH1 and EFH and ZZ domains. The fourth domain is the C-terminal domain (Fig. 12.1). Only a few of these domains have been crystallized and have had their structures solved (Table 12.2). These are the first actin-binding domain ABD1, repeat 1 (R1) and the WW and EFH subdomains in complex with a peptide from β-dystroglycan, an essential partner of dystrophin (see below). Depending on the location of the promotor, the shorter isoforms do not maintain the ABD1 subdomain and do not present all the repeats. As an example, the shortest isoform, Dp71, consists only of the C-terminal end starting at the EFH1 domain.

Utrophin has a similar domain structure to Dystrophin, the only difference being that there are two repeats less than in dystrophin (Winder et al. 1995; Le Rumeur

Table 12.2 Three-dimensional structures of dystrophin, utrophin and spectrin subdomains

Protein	Domain	PDB ID
Dystrophin (H. sapiens)	N-terminal actin binding domain	1DXX
Dystrophin (H. sapiens)	Repeat 1	3UUN
Dystrophin (H. sapiens)	WW, EFH1, EFH2 sub-domains	1EG3, 1EG4
Utrophin (H. sapiens)	N-terminal actin-binding domain	1QAG, 1BHD
Utrophin (R. norvegicus)	Repeat 1	3UUL
Spectrin α (D. melanogaster)	Repeat 14	2SPC
Spectrin α (G. gallus)	Repeats 15, 16, 17	1U4Q
Spectrin α (G. gallus)	Repeats 15, 16	1U5P
Spectrin α (G. gallus)	Repeat 16	1AJ3
Spectrin α (G. gallus)	Repeats 16, 17	1CUN
Spectrin α (G. gallus)	SH3 domain	1SHG, 1 U06
Spectrin α (G. gallus)	Tetramerization site	3F31
Spectrin α (H. sapiens)	Tetramerization site	1OWA
Spectrin αβ (H. sapiens)	Tetramerization sites	3LBX
Spectrin β (H. sapiens)	Repeats 8, 9	1S35
Spectrin β (H. sapiens)	Repeats 13, 14, 15	3KBT
Spectrin β (H. sapiens)	Repeats 14, 15	3F57, 3EDU
Spectrin β (H. sapiens)	Repeats 14, 15, 16	3EDV
Spectrin β (H. sapiens)	CH2 domain	1AA2, 1BKR
Spectrin β (H. sapiens)	PH domain	1WJM
Spectrin β (M. musculus)	PH domain	1MPH

et al. 2010). Apart from the ABD1 domain, which is also found in many other proteins, the other domains show high specificity in terms of their structure and interactions (see below).

Spectrin is a large, mostly heterotetrameric protein with a characteristic modular structure of helical spectrin repeats (Fig. 12.1). While there are 20 full-length repeating units and one incomplete repeat in α spectrins, only 16 (or 29 in case of heavy isoforms) repeats, along with one incomplete repeat, occur in β spectrins. The incomplete segments together provide one full spectrin repeating unit after tetramer formation (Kotula et al. 1993; Ipsaro et al. 2010). However, spectrin is much more than just an assembly of spectrin repeats (Machnicka et al. 2014). The 280-kDa α-spectrin also contains EF hands (calmodulin-like domains) at its N-terminus, an SH3 domain within repeat 9, and a short insert (named CCC) within repeat 10 of αII, which includes sites for calmodulin binding and caspase/calpain cleavage (Czogalla and Sikorski 2005). The variety of β spectrins is reflected not only in a broad molecular weight range (246–430 kDa), but also by the presence or absence of C-terminal plecstrin homology (PH) domains. Moreover, an actin binding domain, which consists of tandem calponin homology (CH) domains, is located at the N-terminus of each of the β spectrins.

12.2.1   Common Features of Dystrophin and Spectrin

12.2.1.1   Classical ABD Domains at the N-terminal End of Dystrophin, Utrophin and β-spectrin

Dystrophin, utrophin and β-spectrins are F-actin interacting proteins through the presence of actin-binding-domains either constituted by two CH domains (respectively CH1 and CH2) situated at the N-terminal end of the mentioned proteins (ABD1 in the case of dystrophin) or constituted of spectrin repeats (ABD2) in dystrophin only.

The entire ABD1 domains of dystrophin and utrophin have been crystallized as open conformations stabilized as head-to-tail homodimers (Keep et al. 1999; Norwood et al. 2000). The relative orientation between the CH1 and CH2 subdomains is different in the dystrophin and utrophin structures and it has been proposed that this is associated with the modulation in the sequence of the linker region bridging the CH subdomains. This linker region is folded into an α-helix which could be bent (Fig. 12.2a), and its amino-acid composition (Fig. 12.2b) influences the thermodynamic stability of the whole subdomain, probably through the conformational equilibrium between the open and the closed states of the ABD in solution (Bandi et al. 2015). The solution conformation adopted by the ABD or its conformation after binding to F-actin is still controversial and may vary between dystrophin and utrophin (Sutherland-Smith et al. 2003; Lin et al. 2011; Broderick et al. 2012; Singh and Mallela 2012). Three actin binding sites (ABS) have been distinguished by experimental mapping on the dystrophin ABD1. This mapping indicated that three independent ABS regions are involved in the contact with F-actin. These are from Lys18 to Ala27 (ABS1), from Val89 to Leu116 (ABS2) and from Leu131 to Ser147 (ABS3). These sites mainly involve the CH1 subdomain (residue 9–121) and the N-terminal end of CH2 (Henderson et al. 2010).

The actin binding domain of β spectrin also consists in two CH subdomains, CH1 (residues 51–156) and CH2 (171–282). The CH2 domain was initially supposed to have a key role in the interactions with actin and this is the reason why only the structure of the CH2 domain was determined at high resolution (1.1 Å; PDB ID: 1BKR), but it appears that isolated CH2 domains bind rather weakly to actin as well (Banuelos et al. 1998). Like the corresponding subdomain in dystrophin and utrophin, CH2 is composed of four parallel helices interconnected by longer loops together with three additional helices resulting in a tight and globular structure. The most conservative elements across the various CH domains are the residues forming helices A, C and G, which play a major role in actin binding. Further on in this Chapter, we discuss a phospholipid-sensitive mechanism of actin binding by this domain.

```
TSATEFENVGNQPPYSRINARWDAPDDELDNDNSSARLFERSRIKALADEREVVQKKTFT
-------------------MLWWEEVED--------------------CYEREDVQKKTFT
-------------------MAKYGEHEASPDNGQNEFSDIIKSR----SDEHNDVQKKTFT
                 .:     :                         . *.: *******

KWVNSHLARVSCR-ITDLYKDLRDGRMLIKLLEVLSGEMLPKPTKGKMRIHCLENVDKAL
KWVNAQFSKFGKQHIENLFSDLQDGRRLLDLLEGLTGQKLPK-EKGSTRVHALNNVNKAL
KWINARFSKSGKPPINDMFTDLKDGRKLLDLLEGLTGTSLPK-ERGSTRVHALNNVNRVL
**:*:.::. .    *  :::.**.*** *:.*** *:* *** .*. *:*.*:**:..*

QFLKEQRVHLENMGSHDIVDGNHRLVLGLIWTIILRFQIQDIVVQTQEGRETRSAKDALL
RVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNIILHWQVKNVMKNIMAGLQQTNSEKILL
QVLHQNNVELVNIGGTDIVDGNHKLTLGLLWSIILHWQVKDVMKDVMSDLQQTNSEKILL
..*:::.* *  *:*. ******.*.***:*.***.:*:::::  :   . :  .::. **

LWCQMKTAGYPHVNVTNFTSSWKDGLAFNALIHKHRPDLIDFDKLK-DSNARHNLEHAFN
SWVRQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAFN
SWVRQTTRPYSQVNVLNFTTSWTDGLAFNAVLHRHKPDLFSWDKVV-KMSPIERLEHAFS
*  . .*  *.:*** ***:**.****;**::* *.***.:::.:  . ..  .*****.

VAERQLGIIPLLDPEDVFTENPDEKSIITYVVAFYH---
IARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLP
KAQTYLGIEKLLDPEDVAVQLPDKKSIIMY---------
 *    *** *******  **:***: *
```

Fig. 12.2 Experimental structures of the actin-binding domain composed of two calponin homology (CH) domains obtained for human dystrophin and utrophin and the two CH2 structures of the β spectrin available in the Protein Data Bank (PDB). (**a**) Backbone representation of the ABD domain emphasizing that the two CH subdomains are mainly composed of alpha-helices. Tridimensional superimposition allows us to establish the very high degree of structural conservation of this domain. (**b**) Primary sequence alignment (spectrin B, dystrophin, utrophin) shows the highest modulations to be located at both terminal ends and at the linker region bridging the two CH subdomains

12.2.1.2 Central Domains are Composed of Repeats with Similar Folds in Triple Helical Coiled-Coils

The common feature of these three proteins is the presence of numerous so-called spectrin repeats made of about 110 residues each and folded in triple helical coiled-coils. These coiled-coils are rather unusual since they are not formed by dimerization as in, for example, myosin or tropomyosin. In this case three successive helices (A, B and C) interrupted by two loops fold back on themselves to produce a three strand coiled coil. The interaction between the three alpha-helices is due to the presence of the classical heptad pattern of residues along the sequence (Winder et al. 1995). The heptad is defined as seven residues labelled "a" to "g" where the residues in position "a" and "d" are preferentially hydrophobic while the other are mostly hydrophilic. This pattern allows the folding in coiled-coils with all the "a" and "d" residues constituting the core of the molecule (Fig. 12.3a).

Spectrin repeats are all rather similar in length being constituted by roughly 106 residues each, but there are variations between 99 and 114 amino acid residues. Unlike spectrins, the related dystrophin repeats are largely heterogeneous in length because of variations in length of their alpha-helices. As an example, repeat 14 composed of 94 amino acids (aa) only has a very short A helix (22 residues) and repeat 10 (92 aa) has a very short C helix (25 residues).

Although the sequence homology between the spectrin repeats does not go beyond 30 % (Leluk et al. 2001), and dystrophin and utrophin share a homology of less than 20 % (Nicolas et al. 2014), all spectrin repeats, as well as dystrophin and utrophin repeats, share a common fold roughly 50 Å in length and 20 Å in diameter. This was first described for the α spectrin of *D. melanogaster* (PDB ID: 2SPC; Yan et al. 1993; Pascual et al. 1997; Muthu et al. 2012; Fig. 12.3b). In a typical spectrin left-handed supercoil segment, helix B is kinked and is slightly longer than the other two helices in the segment (PDB IDs: 1CUN, 1U5P, 1U4Q; Grum et al. 1999; Kusunoki et al. 2004b). However, the positioning of loops interconnecting the helices of individual repeats may vary to some extent (Grum et al. 1999; Kusunoki et al. 2004a). It seems that conservative tryptophan residues are critical for stabilization of the triple-helical structure, although the core of some of the spectrin repeats may deviate from this pattern, as was described for R9 of human β spectrin, where water-mediated hydrogen bonding forms the core (PDB ID: 1S35; Kusunoki et al. 2004a). The tiny dissimilarities among spectrin repeats result in differences in their stabilities. Individual spectrin repeats also have different folding characteristics, as shown for repeats 15–17 of α spectrin (Kwa et al. 2014). Remarkably, helix C and helix A of the next repeat exhibit an uninterrupted helical structure (Fig. 12.3c), thus enabling cooperative propagation of forced unfolding throughout the whole molecule (Kusunoki et al. 2004b; Law et al. 2003). These features, together with a broad array of possible inter-repeat interactions and potential rearrangement of helices within a single repeat, may be keys to understanding the remarkable flexibility of spectrins (Grum et al. 1999; Mirijanian et al. 2007). It is worth underlining that upon full extension the spectrin tetramer is 180–200 nm long, but its physiological rest length is only 55–60 nm. Most recently, a model has been proposed where
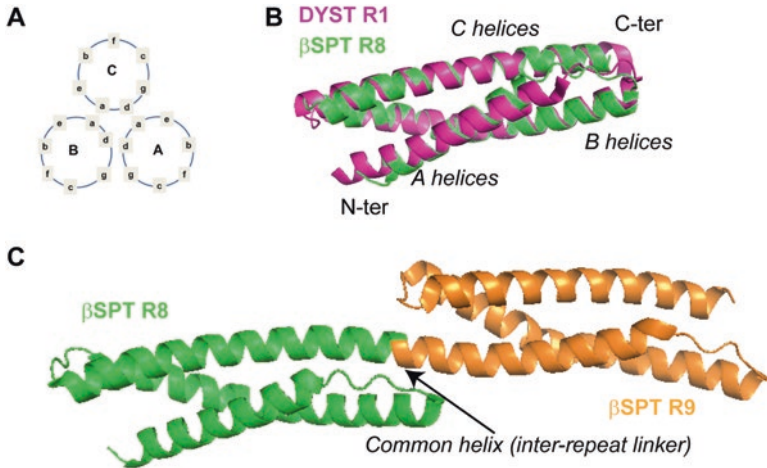
Fig. 12.3 The heptad pattern of the residues in the sequence allows the 3D folding in a triple helical coiled-coil for spectrin and dystrophin repeats. (**a**) Representation as a helix wheel of the coiled-coil structure: the "a" and "d" residues are situated in the core of the molecule while the other residues are accessible to the solvent and for interactions with partners. A, B and C are the three helices of the triple coiled-coil. (**b**) The crystal structures of repeat 8 of β-spectrin (1S35) and repeat 1 of dystrophin (3UUN) are superimposed. The rms deviation between the two structures is 1.333 Å. (**c**) The crystal structure of the R8–9 tandem repeat from β-spectrin showing the common helix constituting the inter-repeat linker joining the repeats

spectrin is a hollow cylinder and the pitch of each spectrin repeat increases during extension (Brown et al. 2015). Inter-repeat linkers do not all appear as helical in dystrophin or utrophin (Legrand et al. 2011; Nicolas et al. 2014).

12.2.1.3 EF Domains at the C-Terminal End of Dystrophin, Utrophin and α-spectrin

The structures of the EF hands of dystrophin have been determined along with the WW domain in complex with a Pro-rich motif of β-dystroglycan (Huang et al. 2000), establishing that the β-dystroglycan interaction with dystrophin is achieved through both WW and EFH1 subdomains. The recognition mechanism has been revealed to be highly similar to what is observed for Pro-rich sequences binding to SH3 domains. Very little is known about utrophin EFH subdomains. The EFH1 and EFH2 of utrophin share respectively 65 % and 77 % of identity with the corresponding subdomains in dystrophin, probably indicating a high degree of structure similarity. The C-terminus of α spectrins also exhibits structural similarities to calmodulin. The two EF-hands (EFH) present in αII spectrins bind cooperatively two calcium ions with rather low affinities in the 50–100 μM range, unlike the analogous structures in αI that do not bind Ca^{2+} (Buevich et al. 2004). It seems that EF-hands stabilize interactions of the CH2 domain of the spectrin dimer with actin

by binding the linker that comes between the actin binding domain and the first β-spectrin repeat (Korsgren and Lux 2010). Moreover, EF-hands can bind some proteins of the membrane skeleton, such as protein 4.2, in a calcium- and calmodulin-dependent manner (Korsgren et al. 2010).

12.2.2 Dystrophin and Spectrin Structural Dissimilarities

Dystrophin is a monomeric protein which has been found to be difficult to crystallize even as small fragments. The only experimentally defined structure was obtained for repeat 1 (R1) constrained to be a dimer induced by an S-S bridge between two repeats (Muthu et al. 2012). This structure was found to be similar to spectrin repeats, post-validating the reliability of homology models obtained previously based on spectrin structures (Legrand et al. 2011). However, this structure does not show how several repeats are chained together in the full length molecule, especially whether or not the linker region is helical across consecutive repeats, and it does not define their relative positioning. Because of this, in a recent study, we obtained structural models of tandem or multi-repeat fragments in solution using small angle X-ray scattering (SAXS) and molecular modelling. This work showed that the multi-repeat fragments present well-defined kinks at some inter-repeat linkers that give a tortuous and complex topology to the whole dystrophin filament (Delalande et al. to be published; Fig. 12.4).

These kinks are situated at specific linkers between successive repeats and are not uniformly distributed all along the central domain of dystrophin. Kinks between spectrin repeats are reminiscent of the angle found at the spectrin ankyrin-binding site, one of the determining features of spectrin-ankyrin interactions (see below). Similarly, the angle observed between repeats 16 and 17 shows that region to be one of the key places for the interaction of dystrophin with neuronal nitric oxide synthase (nNOS; see below; Molza et al. 2015).

Utrophin has not yet been studied with such tools, but it is likely that kinks will be present in utrophin as in dystrophin. These kinks would be placed at different sites than in dystrophin since, unlike dystrophin, repeats 10–15 of utrophin do not bind actin (Amann et al. 1999).

In contrast to dystrophin, the functional units of spectrin are tetramers or higher oligomers (Nans et al. 2011). A broad range of oligomeric states of spectrin in erythrocytes reflects the high degree of elasticity and flexibility of the red blood cell membrane. The α/β heterodimer is stabilized mostly by interactions between two C-terminal and two N-terminal repeats of the α and β subunits, respectively (Begg et al. 2000). The next level of organization is achieved by head to head dimer interactions. The key role in spectrin tetramerization is played by the incomplete repeats of both subunits. The roughly 30 residue long N-terminus of α spectrin forms a single helix and the roughly 70 C-terminal residues of β spectrin fold into two helices. These two structural motifs interact to reconstruct a full triple helical structure where hydrophobic interactions are the major players (PDB ID:
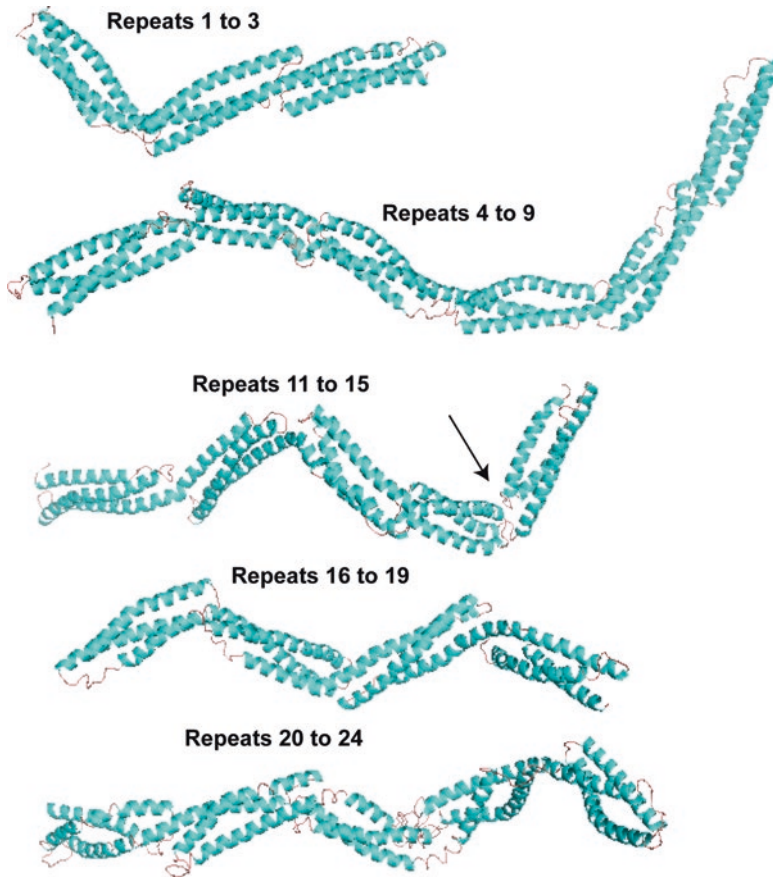
Fig. 12.4 Structural models of multirepeat fragments of dystrophin obtained by small angle X-ray scattering and molecular modeling. Five of the eight fragments studied are shown here. The N-terminal ends are to the left. They all show kinks placed at inter-repeat linkers with angles as high as 90° between repeats 14 and 15 in the R11–15 fragment (*arrow*)

3LBX; Ipsaro et al. 2010). The tetramerization regions of αI and αII (PDB IDs: 1OWA and 3F31, respectively) are not equivalent; the linker following the first helix is either unstructured or helical, respectively (Mehboob et al. 2010). However, after binding β spectrin, the conformation of the linker is helical in both cases. This feature may contribute to the differences in flexibility between various isoforms of spectrin. Although the general model presumes spectrin tetramers or higher oligomers as prerequisite elements of the membrane skeleton, tetramerization of nonerythroid spectrin appears to be unnecessary for normal development of *D. melanogaster* (Khanna et al. 2015).

In all α spectrins, repeat 9 is interrupted by an SH3 domain (Robertsson et al. 2005). The core of this domain consists of antiparallel β strands that form two β-sheets perpendicular to each other (PDB ID: 1SHG; Musacchio et al. 1992).

Recently, more and more data point out the role of the SH3 domain of spectrin in numerous processes connected to cell signalling and repair of DNA interstrand cross-links (for review see Machnicka et al. 2014).

Some isoforms of β spectrin have an extended C-terminus where a PH domain can be found (Macias et al. 1994; Zhang et al. 1995). Seven antiparallel β strands that form a barrel and a subsequent conservative α-helix are the core of such structural motifs, which are known to occur in numerous eukaryotic proteins involved in signal transduction, organization of cytoskeleton and many other cellular processes. Although little is known about the physiological role of the PH domain of spectrin, the positively charged lipid-binding surface seems to be sufficient to bind the inositide head group of PIP_2 (Hyvonen et al. 1995). Complexation to inositol-1,4,5-triphosphate seems to induce ordering of the binding loops when compared to the structure of the free domain in solution (PDB ID: 1MPH; Nilges et al. 1997). On the other hand, it appears that the presence of the PH domain is necessary for membrane targeting of spectrin in midgut copper cells of *D. melanogaster*, but the mechanism is not related to PIP_2- binding activity (Das et al. 2008).

12.3 The Scaffolding Function of Dystrophin and Spectrins: Binding to Phospholipids and Organizing Protein-Protein Assemblies

The first impression when looking at the structures of spectrins and dystrophin is that their repeating units are just modules that build proteins of elongated shape and are predominantly responsible for the observed structural and mechanical properties. In that sense, dystrophin has long been considered as a rod-shape protein linking the cytoskeletal element actin to the extracellular matrix (Koenig et al. 1988; Ervasti and Campbell 1991). However, more and more data show that such modules are also binding sites for a number of cytoskeletal and signal transduction proteins, as well as for phospholipid membranes (Machnicka et al. 2014; Djinovic-Carugo et al. 2002; Boguslawska et al. 2014b; Le Rumeur et al. 2003; Legardinier et al. 2009; Ervasti and Campbell 1993a; Lai et al. 2009; Prins et al. 2009).

12.3.1 Dystrophin and Spectrin: Two Coiled-Coil Filaments Interacting with Membrane Lipids

Phospholipid binding properties of dystrophin are essential for its main biological function to maintain the cohesion of the sarcolemma during compression and re-extension of the membranes of muscle cells. It is well established that a major part of the dystrophin central domain binds to membrane lipids, but with a heterogeneous interaction pattern all along the filament length (Fig. 12.5). The cartography

of the dystrophin central domain emphasizes that, despite a common structural pattern, different repeats show highly different molecular surface properties which define sub-domains involved in specific interactions with various partners (Legrand et al. 2011; Nicolas et al. 2014). For example, repeats R1–R3 strongly interact with anionic lipids, but repeats R20–R24 do not show any similar binding properties (Legardinier et al. 2008). This C-terminal end non-binding feature seems to be a significant difference between dystrophin and spectrin binding properties. Anionic lipids also bind strongly with the subdomain comprising repeats R4–R19 of dystrophin (Legardinier et al. 2009). Another localized binding feature has also been revealed for some repeats of the central domain. For instance, since it binds very weakly to large unilamellar vesicles (LUVs) compared to small unilamellar vesicles (SUVs), except when the LUVs contain PE and/or cholesterol, the binding properties of dystrophin repeat R2 were revealed to be highly sensitive to lipid packing (Le Rumeur et al. 2007). In addition, cholesterol has been shown to increase the insertion of several repeats of dystrophin into the membrane (Ameziane-Le Hir et al. 2014).

For some parts of the central domain, circular dichroism analysis indicated a lower degree of inter-helix interaction in the coiled-coil filament upon binding with phospholipids. This is notably the case in three-repeat fragments like R4–6, R6–9, R12–14 or R17–19. However, coiled-coils should open only partly upon lipid binding, since circular dichroism measurement modulations are not comparable to what can be observed in TFE. Such dynamic behaviour was not observed for longer fragments including R11–15 and R14–17. Among the specific binding profiles encountered along the central domain, the R11–15 fragment seems to be the only subdomain to bind to PC/PE lipids (Legardinier et al. 2009; Sarkis et al. 2011).

Both erythroid and non-erythroid spectrins possess an ability to interact with phospholipids, which has been proven for natural and model membranes (for review see Boguslawska et al. 2014a). Although spectrin repeats exhibit a general lipid binding ability, high affinity sites have recently been recognized (Fig. 12.5). These include PS recognizing repeats 8–10 of the α subunit, and repeats R2–4 and R12–14 of β spectrin (An et al. 2004a, b), as well as a binding site for PE-rich membranes within R14 of β spectrin which partly overlaps with the ankyrin binding domain (described below). Ankyrin-sensitive binding of PE-rich lipids was identified within the N-terminal part of the ankyrin binding domain of both β erythrocytic and β non-erythrocytic spectrin (Hryniewicz-Jankowska et al. 2004; Bok et al. 2007). Further structural studies showed that lipid binding activity is confined to helix C of R14, the highly amphipathic character of which correlates with its mixed 3_{10}/α-helical conformation (Czogalla et al. 2007). Remarkably, interaction with PE/PC membranes or detergents provokes partial opening of the coiled-coil structure of the 14th repeat (Czogalla et al. 2008; Fig. 12.6). The current structural model of the ankyrin-dependent lipid binding site of β spectrin suggests that one of the tryptophan residues within the 3_{10}/α-helix is slightly shifted outside the hydrophobic core of the helix bundle and thus could initiate interactions with membranes (Wolny et al. 2011). Another spectrin region that interacts with phospholipids is the actin binding domain (ABD). Binding of PIP_2 triggers a conformational switch from a closed to
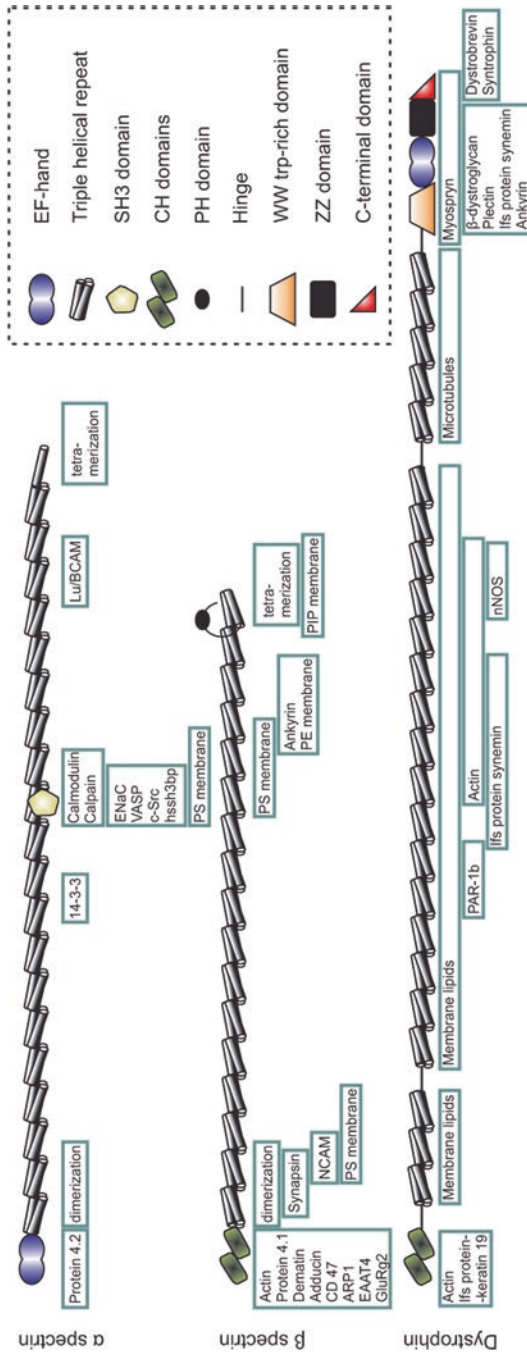
Fig. 12.5 **Interaction partners of spectrin and dystrophin.** Interacting molecules are included in the rectangles with their locations roughly corresponding to sites of interactions

β spectrin segment 1684 - 1896

C-terminus

Membrane

Ankyrin ZU5
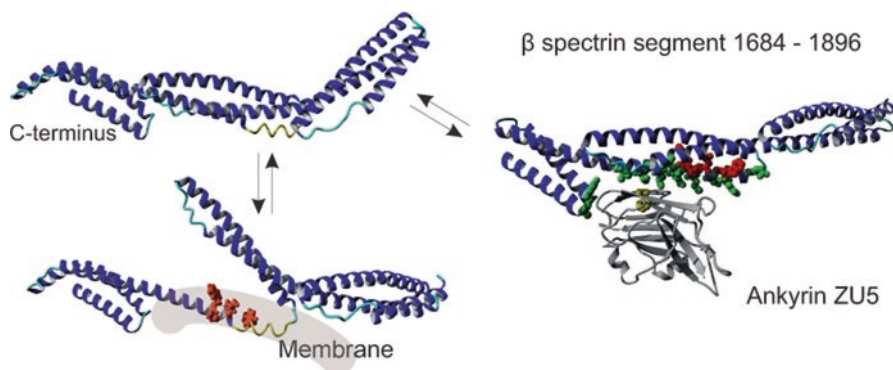
Fig. 12.6 The ankyrin-lipid binding site of spectrin. The triple helical repeats are represented by the ankyrin-sensitive lipid-binding domain within the 14th segment of β-spectrin (ribbon representation of segments 13–15). Membrane binding results in the "open" conformation (amino-acid residues essential for lipid binding in *red*: W1771, L1775, M1778, W1779, all within helix 14C; the 3_{10} helix is *yellow*). Alternatively, steric interference upon the interactions of spectrin with the ZU5 domain of ankyrin (*gray*) blocks the "opening" of the triple helical bundle – PDB ID: 3KBT (amino-acid residues crucial for ankyrin binding in *green*, residue F917 of ZU5 in *yellow*) (Adapted from Boguslawska et al. 2014b)

an open conformation of the CH1-CH2 tandem. This does not influence binding of actin, but enhances interactions with the second target for ABD, namely protein 4.1 (An et al. 2005). On the other hand, structural studies on the CH domains of α-actinin suggest that opening of the domains influences their affinity for actin (Galkin et al. 2010). Out of the various domains of spectrin, the plecstrin homology domains present in some of the isoforms of β subunits seems to be the best candidates for lipid membrane anchors. However, the PH domain of spectrin binds weakly to phosphatidylinosites without specificity towards these lipids, in contrast to the PH domains from some other proteins (Lemmon et al. 2002). Provided that multiple spectrin tetramers form a network containing many lipid binding domains, interactions with membranes may be very strong due to the resulting multivalent character.

12.3.2 Examples of Protein Partners of Dystrophin and Spectrin

The first protein partner of dystrophin to be deswcribed is β-dystroglycan (Campbell and Kahl 1989), which together with dystrophin and other interacting proteins forms the dystrophin-glycoprotein complex (DGC: Fig. 12.7; Ervasti and Campbell 1991), (b). Among muscle membrane proteins dystrophin only interacts with β-dystroglycan, which in turn interacts with intrinsic membrane proteins such as the sarcoglycans. Sarcoglycans and β-dystroglycan interact with proteins of the
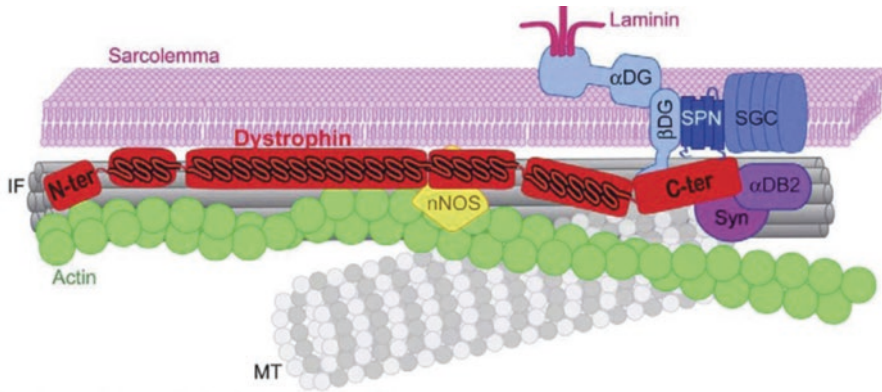
Fig. 12.7 Schematic view of the dystrophin associated complex (DGC). Illustration of the scaffolding role of the dystrophin (in *red*) resulting in the link of sarcolemmal components (in *blue*) to the cytoskeletal protein network. (*MT* microtubule, *IF* intermediate filaments, *DG* dystroglycan, *Syn* syntrophin, *DB* dystrobrevin, *SGC* sarcoglycans)

extra-cellular matrix, including α-dystroglycan and laminins (Ervasti and Campbell 1993a). Intriguingly, dystrophin interacts with three different types of cellular filaments; filamentous actin (Ervasti et al. 1997), intermediate filaments (Bhosle et al. 2006) and microtubules (Prins et al. 2009), which together form a crucial filamentous network for striated muscles. Other proteins involved in signalling or physiology of the muscle or with other unknown functions are also important partners of dystrophin. These include neuronal Nitric Oxide Synthase (nNOS), the polarity associated protein PAR1-b, dystrobrevin, syntrophin and myosprin (Fig. 12.5). The nNOS–dystrophin interaction has been the most intensively studied since it was established that loss of nNOS localization at the sarcolemma was related to an increase in muscle weakness or damage (Sander et al. 2000; Sato et al. 2008) and to a more severe phenotype in patients with mutated dystrophin (Gentil et al. 2012). The primary nNOS binding site was first localized in the dystrophin central domain, involving both repeats R16 and R17 (Lai et al. 2009, 2013). Work combining biochemical experiments, structural data and molecular modelling revealed the association mode of the nNOS PDZ subdomain with the R16-R17 repeats (Molza et al. 2014, 2015). Since PDZ subdomains form classical homodimers, the non-canonical association of PDZ with a three-helix coiled-coil structure seems at first to be similar to the structural organization of the spectrin-ankyrin complex (a beta sheet lying on a three-helix bundle filament bearing a kink; Fig. 12.8). However, detailed analysis of the two structures reveal as many similarities as differerences. The kink in the coiled-coil axis observed at the linker region between successive repeats involved in the interaction with the partner is present in both structures and seems essential for the stabilization of the association. But topologically, as ankyrin binds the spectrin filament at the linker region between two repeats, its contact with the beta sheet is mainly achieved with the end of the helix C of the first repeat and with the loop bridging helices B and C of the second repeat. In contrast, nNOS-PDZ binds
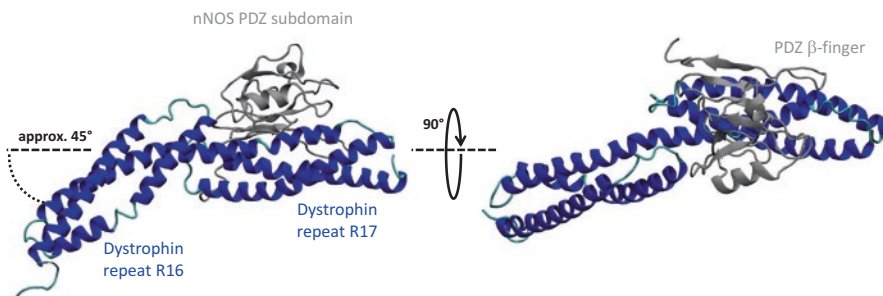
Fig. 12.8 Structural model for nNOS-PDZ–dystrophin R16–17 complex based on biochemical mapping of the interface. Three-helix coiled-coil repeats R16 and R17 of dystrophin are shown in backbone representation colored in *blue* and the nNOS-PDZ subdomain is shown as a *grey* backbone representation. Principal features established to be involved in the specificity of the interaction are highlighted: the filament kink between successive dystrophin repeats and the β-finger contact with the R17 helix C

dystrophin on the other side of the linker region, namely with the begining of the helix A of the second repeat and with the loop bridging helices A and B of the first repeat. In addition, whereas a large extended hydrophobic patch is involved at the nNOS recognition interface of dystrophin (Giudice et al. 2013), specific recognition of ankyrin by spectrin is mainly driven through electrostatics (Ipsaro and Mondragon 2010).

Apparently, the spectrin repeats evolved to gain functional specialization within the conserved three-dimensional fold. Here, the most thoroughly explored example is the ankyrin-binding site which includes highly conserved regions within the 14th and 15th repeats of β spectrin (Czogalla and Sikorski 2010). Moreover, this region also includes an ankyrin-sensitive binding site for PE-rich membranes (Wolny et al. 2011). All these features are hidden in a structurally common tandem of triple-helical spectrin repeats, which was confirmed both in crystal structures (PDB IDs: 3EDU, 3F57, 3EDV; Stabach et al. 2009; Ipsaro et al. 2009; Davis et al. 2009) and in solution (Czogalla et al. 2007). The ankyrin-docking site is formed by conservative anionic amino acid residues on one of the facets of helix C of the 14th repeat and some residues within the inter-repeat linker and the loop between helices B and C of the 15th repeat (Fig. 12.6). Formation of the ankyrin-spectrin complex requires shape complementarity rather than induced fit (Ipsaro et al. 2009), and further studies indicated that the inter-repeat kink is one of the determinants of spectrin-ankyrin interactions (LA-Borde et al. 2010). On the other hand, sensitivity of the spectrin repeats to mechanical stress might be transduced into alterations in ligand (e.g. ankyrin) binding, which could provide selective plasticity of the membrane scaffold and a mechanical switch that links membrane deformations and spatial organization of receptors or signal-transducing molecules (Stabach et al. 2009). The ankyrin-dependent lipid-binding site may be involved in preventing the aggregation of the spectrin tetramers that are linked to some transmembrane proteins via ankyrin. As the number of ankyrin molecules and spectrin tetramers in erythrocytes is similar

and each spectrin tetramer contains two ankyrin/PE binding sites, it seems that the lipid-binding activity of these sites may play a key role in the appropriate spatial arrangement of individual ankyrin-transmembrane complexes (Chorzalska et al. 2010).

12.4   Dystrophin and Spectrin Mutations Related to Pathologies

The two devastating diseases Duchenne and Becker muscular dystrophies (DMD and BMD) occur after mutation of the *DMD* gene (Koenig et al. 1989). The main types of mutation resulting in dystrophinopathies are deletions of one or several exon(s), while there are ~10 % of duplications of exons and ~15 % of single point mutations (Tuffery-Giraud et al. 2009; Flanigan et al. 2009). According to the Monaco rule (Monaco et al. 1988), if the deletion in the *DMD* gene leads to a reading frame shift, there is a total absence of dystrophin in the muscle and severe phenotypes are observed in patients, usually classified as DMD. If the exon deletions do not shift the reading frame, the presence of an internally truncated dystrophin is detected and this is related to a large and variable range in the severity of symptoms as typified by the BMD phenotype. In certain cases, exceptions to the Monaco rule appear where a DMD phenotype with an in-frame mutation is observed. This is particularly prominent when mutations involve the N-terminal actin-binding-domain or the Cys-rich domain affecting the binding of dystrophin to F-actin or β-dystroglycan, respectively.

   DMD is characterized by a progressive muscle weakness involving all skeletal and cardiac muscles and is accompanied by highly elevated creatine kinase blood levels (Kohler et al. 2005; Davies and Nowak 2006; Rahimov and Kunkel 2013). The early clinical signs leading to a DMD diagnosis are severe difficulties to walk and to climb stairs in the second or third year of childhood. DMD young patients are never able to run and are confined to a wheelchair before the age of 12 years (main criteria for DMD diagnosis). Progressively, respiratory and cardiac impairments are observed and DMD patients mainly die in their early thirties (Wein et al. 2015). All these specific DMD clinical signs are similarly observed in BMD patients, nevertheless with a very broad range of time courses and severity. Some BMD patients are considered as asymptomatic, while others become wheelchair confined around 16 years of age. Some BMD patients survive to very old ages while others die prematurely from heart failure (Bushby and Gardner-Medwin 1993; Bushby et al. 1993). Histological analyses of DMD muscles show cycles of fiber necrosis and regeneration. The regeneration process is however overtaken by fiber loss mechanisms and fibrosis, as adipose tissue replacement is also increased. All these features are vary markedly during DMD evolution or in different types of BMD. For instance the fibrosis and fatty infiltration increase with the age in DMD patients. Finally, electron microscopy studies have revealed lesions of the plasma

membrane in the case of DMD indicating that the maintenance of plasma membrane integrity is the essential role of dystrophin in muscle cells (Petrof et al. 1993).

From a structural point of view, the effect of exon deletion occurring in the central domain of dystrophin and associated with the BMD disease depends on the maintainance of the repeat phasing. These deletions disrupt repeat coding, but this results in main two cases. Indeed, after exon deletion which maintains the reading frame a new junction is created in the shortened dystrophin. This new junction either restores a hybrid repeat comprising a three-helix coiled-coil or it constitutes a new structural motif not in a coiled-coil fold, but leading to the formation of a fractional repeat. The first structural evidence for these rules was obtained through the computation of BMD shortened dystrophin fragments (Nicolas et al. 2012) and has recently been confirmed by the experimental determination of the SAXS-based models for a dystrophin fragment bearing the most frequent BMD deletion of exons 45–47 (Δ45–47; Delalande et al. to be published). Analysis of this model based on low-resolution experimental data clearly shows that the structural organization of the coiled-coil filament is strongly disturbed at the newly created junction. It is important to notice that, with the changed exon phasing, this alteration of the structure leads to the loss of the previously characterized binding site of nNOS. In addition, we showed that the two types of structure of the internally deleted dystrophins produced in these BMD patients could partly explain the differences in the clinical severity of the patients (Nicolas et al. 2015).

The most promising strategies for DMD therapy are to transform a DMD patient into a BMD patient with an asymptomatic phenotype by using exon-skipping and gene or cell therapies.

Exon-skipping aims to transform a DMD patient with an out-of-frame deletion into a BMD patient by deleting additional exons to restore the reading frame (Aartsma-Rus and Van Ommen 2007; Goyenvalle et al. 2004). The most valuable deletions may be those restoring a hybrid repeat at the new junction and the studies reporting the correlation between dystrophin structure and clinical outcomes of BMD patients are highly valuable.

Gene therapy aims to deliver DNA sequences expressing the most important parts of the protein into skeletal and heart cells. The *DMD* gene is the largest human gene and it is not possible to put the entire cDNA sequence into a unique vector that can reach all muscle cells. Therefore, the concept of micro-dystrophin has emerged (Harper et al. 2002). Truncated gene coding sequences (micro-dystrophins), inspired by the truncated dystrophin coding sequences observed in mild BMD patients (England et al. 1990), have been designed (Seto et al. 2012; Fairclough et al. 2013; Jarmin et al. 2014; Mcgreevy et al. 2015). These highly simplified micro-dystrophins only consist of the N- and Cys-rich domains with two hinges and several repeats (Gregorevic et al. 2008; Foster et al. 2008). However, they do not replicate all the functions of dystrophin (Seto et al. 2012) and further improvements are needed based on our knowledge about the structure and function of the central domain of dystrophin (Wilton et al. 2015).

The spectrin-based membrane skeleton is a prerequisite for maintaining the integrity and flexibility of cells. This is particularly important for erythrocytes which experience strong mechanical stresses during high speed flow in the circulation and major deformations due to their passage through capillaries of diameter much smaller than the size of the cells. Thus, defects in the components of the spectrin-based skeleton are reflected in the fragility, fragmentation and premature destruction of red blood cells. Such disorders lead to hemolytic anemias, including hereditary spherocytosis (HS), elliptocytosis (HE), pyropoikylocytosis (HPP) and stomatocytosis (HSt). In the case of the former, most of the molecular defects concern membrane skeleton components involved in vertical interactions, namely spectrin, ankyrin, anion exchanger 1 and protein 4.2. On the other hand, HE and HPP are characterized by molecular defects of proteins involved in horizontal interactions within the membrane skeleton, namely spectrin and protein 4.1 (Gallagher 2004b). Very little is known about the molecular background of HSt, a rare red cell disorder resulting from altered intracellular cation content and some cell volume alterations (Bruce 2009; Boguslawska et al. 2010).

Hereditary spherocytosis (HS) is the most widely spread inherited disease linked to the red blood cell membrane, affecting more than one per two thousand individuals (Da Costa et al. 2013). In three-quarters of cases the inheritance is dominant. The decreased deformability and spheroidal shape of erythrocytes is associated with the loss of membrane surface area in relation to intracellular volume (Eber and Lux 2004). Such cells are sequestered in the spleen and subsequently phagocytosed by macrophages, which results in anaemia and splenomegaly. Manifestation of spherocytosis may vary substantially from patient to patient, starting from very few spherocytes in approx. 35 % of mild cases of HS. Minor to moderate forms of the disease are characteristic of defects in the *SPTB* gene. These cases comprise 15–30 % of the HS population and the transmission is autosomal dominant, although sporadic mutations have also been described. More than 25 mutations elucidated in this gene include ten nonsense or non-coding sequences, ten null mutations and five missense mutations (Bolton-Maggs et al. 2012). The latter group is represented by spectrin Atlanta and spectrin Kissimmee, where the mutation is located in the conserved region that plays a role in interactions with protein 4.1. More severe forms of HS have their sources in mutations in the *SPTA1* gene, although these are relatively rare (<5 % of HS population). In these cases, homozygous or compound heterozygous mutations lead to the development of the disease and can be related to the fact that in erythroid cells α-spectrin is produced in large excess over β-spectrin. An interesting example is the low expression allele α^{LEPRA} (low expression allele of *SPTA1* gene, 3.3 %) that contains a C-to-T transition at position −99 of intron 30. The mutation activates an alternative acceptor site 70 nucleotides upstream from the usual site, which leads to mis-splicing and a frameshift in pre-mRNA. α^{LEPRA} is supposed to be broadly involved in recessive or non-dominant HS, although it is silent in normal individuals. Severe hemolytic anaemia is a result of its combination in *trans* with the α-spectrin Prague mutation (Wichterle et al. 1996). The Prague mutation leads to exon skipping, frameshifts and the production of a truncated protein.

Hereditary elliptocytosis (HE) is another common disease related to mutations in genes encoding membrane skeleton proteins. In general, HE is distributed worldwide, ranging from 1 to 2 cases per ten thousand Caucasians and up to 2 % of the population in malaria endemic regions (Glele-Kakai et al. 1996; Dhermy et al. 2007). HE individuals have a resistance to the parasite *P.falciparum*, as the absence of protein 4.1 and mutations in the spectrin tetramerization site impede binding of the parasite to the host cells (Waller et al. 2003). The hallmark of HE is a mechanically unstable cell membrane, mostly due to mutations in spectrins (90 % of cases). The shape of erythrocytes becomes elliptic, although the cell morphology and clinical severity is heterogeneous. While the majority of HE cases are asymptomatic, homozygous and compound heterozygous individuals suffer from mild-to-severe anaemia, jaundice and splenomegaly (Gallagher 2004a). Moreover, a few cases of fatal hydrops fetalis were reported (Gallagher et al. 1995). Greater membrane fragility, reflected by the presence of numerous poikilocytes, together with thermal sensitivity of erythrocytes, was initially considered as a distinct type of disorder called hereditary pyropoikylocytosis (HPP), but in fact it corresponds to severe cases of HE (Lecomte et al. 1987). The vast majority of HE cases are related to the disturbed ability of the spectrin dimers to self-aggregate into functional tetramers, which are crucial for mechanical stability of the red blood cell membrane (Lecomte et al. 1993). Indeed, most of the defects are linked to missense mutations located at or in the vicinity of the spectrin tetramerization sites; in the N-terminal helix C and first spectrin repeat of the α subunit or within the two C-terminal helices of the β subunit. Some mutations lead to truncations of the C-terminus of β spectrin. In general, location of the molecular defects determines the clinical phenotype of the disease (Lecomte et al. 1993). When mutations are present within the helices critical to the formation of the spectrin tetramer (Nicolas et al. 1998), and they involve amino acid residues that play a key role in stabilization of the spectrin repeat restored after self-association (Zhang et al. 2001; Gaetani et al. 2008), patients experience a more severe form of HE. It is worth underlining that more than 25 mutations related to HE are located within the *SPTA1* gene, while the *SPTB* gene is less affected. This is related to the issue that another factor that determines the severity of HE is the existence of a low expression α allele in *trans*. For example, the α[LELY] (Low Expression Lyon) allele has a C-to-T mutation at nucleotide −12 of intron 45 which leads to deletion of a few amino acid residues within the 20th repeat of α-spectrin. As a consequence, αβ dimer formation is compromised (Wilmotte et al. 1997).

The fact that up to 3 % of the total protein in brain is represented by spectrin reflects how important this protein is in neuronal cells where it interacts with numerous membrane proteins (Davis and Bennett 1983; Machnicka et al. 2014). In neuronal axons spectrin forms a regular network of spectrin tetramers connected to junctional complexes (Xu et al. 2013). Of particular importance is the involvement of non-erythrocytic β spectrins in the docking of synaptic vesicles to the presynaptic membrane, mostly through interactions with synapsin (Sikorski et al. 1991). The latter binds to a roughly 25 amino acid residue segment of the β subunit, and this process is critical for neurotransmission (Sikorski et al. 2000; Zimmer et al. 2000). Deletion mutations of α or β spectrin in *D. melanogaster* result in a severe distortion

of synaptic transmission and disrupt the spatial arrangement of synaptic compo-
nents (Featherstone et al. 2001).

West syndrome is a relatively rare epileptic disease affecting infants. The basis
of the disease apparently relates to distortions in neurotransmitter function. Most
recently it was shown that some mutations in αII-spectrin are related to the West
syndrome (Saitsu et al. 2010; Writzl et al. 2012). The observed mutations lead to
severe cerebral hypomyelination, decreased white matter, brain atrophy and reduced
corpus callosum. A deletion within the helix linker between repeats 19 and 20 and
duplication within repeat 20 of α-spectrin affect the region of the protein responsi-
ble for the formation of spectrin heterodimers. This may disturb β-spectrin-mediated
stabilization of membrane proteins and axonal transport (Holleran et al. 2001;
Lorenzo et al. 2010). In this context it is also worth mentioning that αII-spectrin
plays an important role in the maintenance of the integrity of myelinated axons, and
is enriched at the paranodes which flank Ranvier's node (Ogawa et al. 2006; Voas
et al. 2007).

Mutations in the *SPTBN2* gene have been connected with spinocerebellar ataxias
(SCAs; Ikeda et al. 2006). Such disorders are characterized by degeneration of cer-
ebellum, spinal cord and brainstem. A 39 nucleotide deletion within exon 12 of the
the gene coding βIII-spectrin, causing deletion of amino acid residues 532–544, was
found in an American family suffering from spinocerebellar ataxia type 5. Similarly,
a 15 nucleotide deletion in exon 14, that leads to deletion of amino acid residues
629–634 and introduction of a Trp residue, was found in a French family. Both
mutations refer to spectrin repeat 3, which plays a role in the dimerization of α and
β subunits. Another example is a T-to-C transition in exon 7 resulting in a substitu-
tion (Leu253Pro) in the highly conserved region of the calponin homology domain.
βII-spectrin is mainly expressed in Purkinje cells, where it plays a role in the stabi-
lization of membrane proteins including the glutamate transporter EAAT4 (Ohara
et al. 1998; Jackson et al. 2001). Mutations of spectrin prevent proper localization
of EAAT4 and GluRg2 on the plasma membrane. Loss of these proteins at the cell
membrane may lead to glutamate signalling abnormalities that provoke Purkinie
cell death and eventually lead to ataxia. βII-spectrin is also associated with Golgi
and vesicles (Stankewich et al. 1998) and binds to dynactin, which may suggest a
role of spectrin in vesicular transport (Holleran et al. 2001). Mutations in the CH
domain of βII-spectrin may alter its ability to bind to the actin skeleton and thus
affect the intracellular transport and/or stabilization of membrane proteins.

12.5   Conclusion

As they share a common structural organization – i.e. a multi-domain filament that
includes actin-binding, lipid-binding and EF subdomains and a succession of
coiled-coil repeats forming a central domain – Dystrophin and Spectrin are two
proteins presenting highly similar topologies and interaction networks. All these
specificities are closely related to their scaffolding role, which has been revealed to

be essential for the maintenance of the architecture of the cell. Nevertheless, in contrast to spectrin, dystrophin is coded by a single gene, it is non-ubiquitous and monomeric in solution. The comparison of these two proteins is also limited by the focus on the very different pathologies that emerge from the mutation of their encoding genes. Recent structural studies dedicated to the filamentous coiled-coil central domain of both dystrophin and spectrin have provided meaningful details in order to highlight their specific role in the cell and are promising in the pursuit towards understanding the organization of the large macromolecular assemblies managed by these two dissimilar sisters of the same family.

References

Aartsma-Rus A, Van Ommen GJ (2007) Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. RNA 13:1609–1624

Amann KJ, Guo AWX, Ervasti JM (1999) Utrophin lacks the rod domain actin binding activity of dystrophin. J Biol Chem 274:35375–35380

Ameziane-Le Hir S, Raguénès-Nicol C, Paboeuf G, Nicolas A, Chéron A, Le Rumeur E, Vié V (2014) Cholesterol favors higher level of insertion and organization of spectrin-like repeat 16–21 of human dystrophin in membrane. Biochim Biophys Acta 1838:1266–1273

An X, Guo X, Sum H, Morrow J, Gratzer W, Mohandas N (2004a) Phosphatidylserine binding sites in erythroid spectrin: location and implications for membrane stability. Biochemistry 43:310–315

An X, Guo X, Wu Y, Mohandas N (2004b) Phosphatidylserine binding sites in red cell spectrin. Blood Cells Mol Dis 32:430–432

An X, Debnath G, Guo X, Liu S, Lux SE, Baines A, Gratzer W, Mohandas N (2005) Identification and functional characterization of protein 4.1R and actin-binding sites in erythrocyte beta spectrin: regulation of the interactions by phosphatidylinositol-4,5-bisphosphate. Biochemistry 44:10681–10688

Bandi S, Singh SM, Mallela KM (2015) Interdomain linker determines primarily the structural stability of dystrophin and utrophin tandem calponin-homology domains rather than their actin-binding affinity. Biochemistry 54:5480–5488

Banuelos S, Saraste M, Djinovic Carugo K (1998) Structural comparisons of calponin homology domains: implications for actin binding. Structure 6:1419–1431

Begg GE, Harper SL, Morris MB, Speicher DW (2000) Initiation of spectrin dimerization involves complementary electrostatic interactions between paired triple-helical bundles. J Biol Chem 275:3279–3287

Berghs S, Aggujaro D, Dirkx R JR., Maksimova E, Stabach P, Hermel JM, Zhang JP, Philbrick W, Slepnev V, Ort T & Solimena M (2000) betaIV spectrin, a new spectrin localized at axon initial segments and nodes of ranvier in the central and peripheral nervous system. J Cell Biol, 151, 985–1002.

Bhosle RC, Michele DE, Campbell KP, Li Z, Robson RM (2006) Interactions of intermediate filament protein synemin with dystrophin and utrophin. Biochem Biophys Res Commun 346:768–777

Boguslawska DM, Machnicka B, Sikorski AF (2010) Hereditary stomatocytoses--diagnostic problems and their molecular basis. Pol Merkur Lekarski 29:119–124

Boguslawska DM, Heger E, Listowski M, Wasinski D, Kuliczkowski K, Machnicka B, Sikorski AF (2014a) A novel L1340P mutation in the ANK1 gene is associated with hereditary spherocytosis? Br J Haematol 167:269–271

Boguslawska DM, Machnicka B, Hryniewicz-Jankowska A, Czogalla A (2014b) Spectrin and phospholipids – the current picture of their fascinating interplay. Cell Mol Biol Lett 19:158–179

Bok E, Plazuk E, Hryniewicz-Jankowska A, Chorzalska A, Szmaj A, Dubielecka PM, Stebelska K, Diakowski W, Lisowski M, Langner M, Sikorski AF (2007) Lipid-binding role of betaII-spectrin ankyrin-binding domain. Cell Biol Int 31:1482–1494

Bolton-Maggs PH, Langer JC, Iolascon A, Tittensor P, King MJ (2012) Guidelines for the diagnosis and management of hereditary spherocytosis–2011 update. Br J Haematol 156:37–49

Broderick MJ, Bobkov A, Winder SJ (2012) Utrophin ABD binds to F-actin in an open conformation. FEBS Open Bio 2:6–11

Brown JW, Bullitt E, Sriswasdi S, Harper S, Speicher DW, Mcknight CJ (2015) The physiological molecular shape of spectrin: a compact supercoil resembling a chinese finger trap. PLoS Comput Biol 11:e1004302

Bruce LJ (2009) Hereditary stomatocytosis and cation-leaky red cells--recent developments. Blood Cells Mol Dis 42:216–222

Buevich AV, Lundberg S, Sethson I, Edlund U, Backman L (2004) NMR studies of calcium-binding to mutant alpha-spectrin EF-hands. Cell Mol Biol Lett 9:167–186

Bushby KM, Gardner-Medwin D (1993) The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy I natural history. J Neurol 240:98–104

Bushby KM, Gardner-Medwin D, Nicholson LV, Johnson MA, Haggerty ID, Cleghorn NJ, Harris JB, Bhattacharya SS (1993) The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy II correlation of phenotype with genetic and protein abnormalities. J Neurol 240:105–112

Byers TJ, Lidov HG, Kunkel LM (1993) An alternative dystrophin transcript specific to peripheral nerve. Nat Genet 4:77–81

Campbell K, Kahl S (1989) Association of dystrophin and an integral membrane glycoprotein. Nature 338:259–262

Chelly J, Hamard G, Koulakoff A, Kaplan JC, Kahn A, Berwald-Netter Y (1990) Dystrophin gene transcribed from different promoters in neuronal and glial cells. Nature 344:64–65

Cianci CD, Zhang Z, Pradhan D, Morrow JS (1999) Brain and muscle express a unique alternative transcript of alphaII spectrin. Biochemistry 38:15721–15730

Czogalla A, Sikorski AF (2005) Spectrin and calpain: a 'target' and a 'sniper' in the pathology of neuronal cells. Cell Mol Life Sci 62:1913–1924

Czogalla A, Sikorski AF (2010) Do we already know how spectrin attracts ankyrin? Cell Mol Life Sci 67:2679–2683

Czogalla A, Jaszewski AR, Diakowski W, Bok E, Jezierski A, Sikorski AF (2007) Structural insight into an ankyrin-sensitive lipid-binding site of erythroid beta-spectrin. Mol Membr Biol 24:215–224

Czogalla A, Grzymajlo K, Jezierski A, Sikorski AF (2008) Phospholipid-induced structural changes to an erythroid beta spectrin ankyrin-dependent lipid-binding site. Biochim Biophys Acta 1778:2612–2620

Chorzalska A, Lach A, Borowik T, Wolny M, Hryniewicz-Jankowska A, Kolondra A, Langner M, Sikorski AF (2010) The effect of the lipid-binding site of the ankyrin-binding domain of erythroid beta-spectrin on the properties of natural membranes and skeletal structures. Cell Mol Biol Lett 15:406–423

D'souza VN, Nguyen TM, Morris GE, Karges W, Pillers DA, Ray PN (1995) A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet 4:837–842

Da Costa L, Galimand J, Fenneteau O, Mohandas N (2013) Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders. Blood Rev 27:167–178

Das A, Base C, Manna D, Cho W, Dubreuil RR (2008) Unexpected complexity in the mechanisms that target assembly of the spectrin cytoskeleton. J Biol Chem 283:12643–12653

Davies KE, Nowak KJ (2006) Molecular mechanisms of muscular dystrophies: old and new players. Nat Rev Mol Cell Biol 7:762–773

Davis J, Bennett V (1983) Brain spectrin Isolation of subunits and formation of hybrids with erythrocyte spectrin subunits. J Biol Chem 258:7757–7766

Davis L, Abdi K, Machius M, Brautigam C, Tomchick DR, Bennett V, Michaely P (2009) Localization and structure of the ankyrin-binding site on beta2-spectrin. J Biol Chem 284:6982–6987

Dhermy D, Schrevel J, Lecomte MC (2007) Spectrin-based skeleton in red blood cells and malaria. Curr Opin Hematol 14:198–202

Djinovic-Carugo K, Gautel M, Ylanne J, Young P (2002) The spectrin repeat: a structural platform for cytoskeletal protein assemblies. FEBS Lett 513:119–123

Dubreuil RR, Grushko T (1998) Genetic studies of spectrin: new life for a ghost protein. BioEssays 20:875–878

Eber S, Lux SE (2004) Hereditary spherocytosis – defects in proteins that connect the membrane skeleton to the lipid bilayer. Semin Hematol 41:118–141

England S, Nicholson L, Johnson M, Forrest S, Love D, Zubrzycka-Gaarn E, Bulman D, Harris J, Davies K (1990) Very mild muscular dystrophy associated with the deletion of 46 % of the dystrophin. Nature 343:180–182

Ervasti J, Campbell K (1991) Membrane organization of the dystrophin-glycoprotein complex. Cell 66:1121–1131

Ervasti J, Campbell K (1993a) Dystrophin and the membrane skeleton. Curr Opin Cell Biol 5:82–87

Ervasti J, Campbell K (1993b) A Role for the dystrophin-glycoprotein complex as aTransmembrane linker between laminin and actin. J Cell Biol 122:809–823

Ervasti JM, Rybakova IN, Amann KJ (1997) A multiple site side binding model for the interaction of dystrophin with F-actin. Soc Gen Physiol Ser 52:31–44

Fairclough RJ, Wood MJ, Davies KE (2013) Therapy for duchenne muscular dystrophy: renewed optimism from genetic approaches. Nat Rev Genet 14:373–378

Featherstone DE, Davis WS, Dubreuil RR, Broadie K (2001) Drosophila alpha- and beta-spectrin mutations disrupt presynaptic neurotransmitter release. J Neurosci 21:4215–4224

Flanigan KM, Dunn DM, Von Niederhausern A, Soltanzadeh P, Gappmaier E, Howard MT, Sampson JB, Mendell JR, Wall C, King WM, Pestronk A, Florence JM, Connolly AM, Mathews KD, Stephan CM, Laubenthal KS, Wong BL, Morehart PJ, Meyer A, Finkel RS, Bonnemann CG, Medne L, Day JW, Dalton JC, Margolis MK, Hinton VJ, Weiss RB (2009) Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat 30:1657–1666

Foster H, Sharp PS, Athanasopoulos T, Trollet C, Graham IR, Foster K, Wells DJ, Dickson G (2008) Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer. Mol Ther 16:1825–1832

Gaetani M, Mootien S, Harper S, Gallagher PG, Speicher DW (2008) Structural and functional effects of hereditary hemolytic anemia-associated point mutations in the alpha spectrin tetramer site. Blood 111:5712–5720

Galkin VE, Orlova A, Salmazo A, Djinovic-Carugo K, Egelman EH (2010) Opening of tandem calponin homology domains regulates their affinity for F-actin. Nat Struct Mol Biol 17:614–616

Gallagher PG (2004a) Hereditary elliptocytosis: spectrin and protein 4.1R. Semin Hematol 41:142–164

Gallagher PG (2004b) Update on the clinical spectrum and genetics of red blood cell membrane disorders. Curr Hematol Rep 3:85–91

Gallagher PG, Weed SA, Tse WT, Benoit L, Morrow JS, Marchesi SL, Mohandas N, Forget BG (1995) Recurrent fatal hydrops fetalis associated with a nucleotide substitution in the erythrocyte beta-spectrin gene. J Clin Invest 95:1174–1182

Gentil C, Leturcq F, Ben Yaou R, Kaplan JC, Laforet P, Penisson-Besnier I, Espil-Taris C, Voit T, Garcia L, Pietri-Rouxel F (2012) Variable phenotype of del45-55 Becker patients correlated with nNOSmu mislocalization and RYR1 hypernitrosylation. Hum Mol Genet 21:3449–3460

Giudice E, Molza A-E, Laurin Y, Nicolas A, Le Rumeur E, Delalande O (2013) Molecular clues to the dystrophin–nNOS interaction: a theoretical approach. Biochemistry 52:7777–7784

Glele-Kakai C, Garbarz M, Lecomte MC, Leborgne S, Galand C, Bournier O, Devaux I, Gautero H, Zohoun I, Gallagher PG, Forget BG, Dhermy D (1996) Epidemiological studies of spectrin mutations related to hereditary elliptocytosis and spectrin polymorphisms in Benin. Br J Haematol 95:57–66

Goyenvalle A, Vulin A, Fougerousse F, Leturcq F, Kaplan J, Garcia L, Danos O (2004) Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. Science 306:1796–1799

Gregorevic P, Blankinship MJ, Allen JM, Chamberlain JS (2008) Systemic microdystrophin gene delivery improves skeletal muscle structure and function in old dystrophic mdx mice. Mol Ther 16:657–664

Grum VL, Li D, Macdonald RI, Mondragon A (1999) Structures of two repeats of spectrin suggest models of flexibility. Cell 98:523–535

Harper SQ, Hauser MA, Dellorusso C, Duan D, Crawford RW, Phelps SF, Harper HA, Robinson AS, Engelhardt JF, Brooks SV, Chamberlain JS (2002) Modular flexibility of dystrophin: implications for gene therapy of duchenne muscular dystrophy. Nat Med 8:253–261

Hayes NV, Scott C, Heerkens E, Ohanian V, Maggs AM, Pinder JC, Kordeli E, Baines AJ (2000) Identification of a novel C-terminal variant of beta II spectrin: two isoforms of beta II spectrin have distinct intracellular locations and activities. J Cell Sci 113(Pt 11):2023–2034

Helliwell TR, Man NT, Morris GE, Davies KE (1992) The dystrophin-related protein, utrophin, is expressed on the sarcolemma of regenerating human skeletal muscle fibres in dystrophies and inflammatory myopathies. Neuromuscul Disord 2:177–184

Henderson DM, Lee A, Ervasti JM (2010) Disease-causing missense mutations in actin binding domain 1 of dystrophin induce thermodynamic instability and protein aggregation. Proc Natl Acad Sci U S A 107:9632–9637

Holleran EA, Ligon LA, Tokito M, Stankewich MC, Morrow JS, Holzbaur EL (2001) beta III spectrin binds to the Arp1 subunit of dynactin. J Biol Chem 276:36598–36605

Hryniewicz-Jankowska A, Bok E, Dubielecka P, Chorzalska A, Diakowski W, Jezierski A, Lisowski M, Sikorski AF (2004) Mapping of an ankyrin-sensitive, phosphatidylethanolamine/phosphatidylcholine mono- and bi-layer binding site in erythroid beta-spectrin. Biochem J 382:677–685

Huang X, Poy F, Zhang R, Joachimiak A, Sudol M, Eck MJ (2000) Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. Nat Struct Biol 7:634–638

Hugnot JP, Gilgenkrantz H, Vincent N, Chafey P, Morris GE, Monaco AP, Berwald-Netter Y, Koulakoff A, Kaplan JC, Kahn A et al (1992) Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in nonmuscle tissues. Proc Natl Acad Sci U S A 89:7506–7510

Hyvonen M, Macias MJ, Nilges M, Oschkinat H, Saraste M, Wilmanns M (1995) Structure of the binding site for inositol phosphates in a PH domain. EMBO J 14:4676–4685

Ikeda Y, Dick KA, Weatherspoon MR, Gincel D, Armbrust KR, Dalton JC, Stevanin G, Durr A, Zuhlke C, Burk K, Clark HB, Brice A, Rothstein JD, Schut LJ, Day JW, Ranum LP (2006) Spectrin mutations cause spinocerebellar ataxia type 5. Nat Genet 38:184–190

Ipsaro JJ, Mondragon A (2010) Structural basis for spectrin recognition by ankyrin. Blood 115:4093–5101

Ipsaro JJ, Huang L, Mondragon A (2009) Structures of the spectrin-ankyrin interaction binding domains. Blood 113:5385–5393

Ipsaro JJ, Harper SL, Messick TE, Marmorstein R, Mondragon A, Speicher DW (2010) Crystal structure and functional interpretation of the erythrocyte spectrin tetramerization domain complex. Blood 115:4843–4852

Jackson M, Song W, Liu MY, Jin L, Dykes-Hoberg M, Lin CI, Bowers WJ, Federoff HJ, Sternweis PC, Rothstein JD (2001) Modulation of the neuronal glutamate transporter EAAT4 by two interacting proteins. Nature 410:89–93

Jarmin S, Kymalainen H, Popplewell L, Dickson G (2014) New developments in the use of gene therapy to treat Duchenne muscular dystrophy. Expert Opin Biol Ther 14:209–230

Keep NH, Norwood FLM, Moores CA, Winder SJ, Kendrick-Jones J (1999) The 2.0 a structure of the second calponin homology domain from the actin-binding region of the dystrophin homologue utrophin. J Mol Biol 285:1257–1264

Khanna MR, Mattie FJ, Browder KC, Radyk MD, Crilly SE, Bakerink KJ, Harper SL, Speicher DW, Thomas GH (2015) Spectrin tetramer formation is not required for viable development in Drosophila. J Biol Chem 290:706–715

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50:509–517

Koenig M, Monaco AP, Kunkel LM (1988) The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 53:219–226

Koenig M, Beggs A, Moyer M, Scherpf S, Heindrich K, Bettecken T, Meng G, Muller C, Lindlof M, Kaariainen H (1989) The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. Am J Hum Genet 45:498–506

Kohler M, Clarenbach CF, Boni L, Brack T, Russi EW, Bloch KE (2005) Quality of life, physical disability, and respiratory impairment in Duchenne muscular dystrophy. Am J Respir Crit Care Med 172:1032–1036

Korsgren C, Lux SE (2010) The carboxyterminal EF domain of erythroid alpha-spectrin is necessary for optimal spectrin-actin binding. Blood 116:2600–2607

Korsgren C, Peters LL, Lux SE (2010) Protein 4.2 binds to the carboxyl-terminal EF-hands of erythroid alpha-spectrin in a calcium- and calmodulin-dependent manner. J Biol Chem 285:4757–4770

Kotula L, Desilva TM, Speicher DW, Curtis PJ (1993) Functional characterization of recombinant human red cell alpha-spectrin polypeptides containing the tetramer binding site. J Biol Chem 268:14788–14793

Kusunoki H, Macdonald R, Mondragon A (2004a) Structural insights onto the stability and flexibility of unusual erythroid spectrin repeats. Structure 12:645–656

Kusunoki H, Minasov G, Macdonald R, Mondragon A (2004b) Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin. J Mol Biol 344:495–511

Kwa LG, Wensley BG, Alexander CG, Browning SJ, Lichman BR, Clarke J (2014) The folding of a family of three-helix bundle proteins: spectrin R15 has a robust folding nucleus, unlike its homologous neighbours. J Mol Biol 426:1600–1610

La-Borde PJ, Stabach PR, Simonovic I, Morrow JS, Simonovic M (2010) Ankyrin recognizes both surface character and shape of the 14-15 di-repeat of beta-spectrin. Biochem Biophys Res Commun 392:490–494

Lai Y, Thomas GD, Yue Y, Yang HT, Li D, Long C, Judge L, Bostick B, Chamberlain JS, Terjung RL, Duan D (2009) Dystrophins carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy. J Clin Invest 119:624–635

Lai Y, Zhao J, Yue Y, Duan D (2013) alpha2 and alpha3 helices of dystrophin R16 and R17 frame a microdomain in the alpha1 helix of dystrophin R17 for neuronal NOS binding. Proc Natl Acad Sci U S A 110:525–530

Law R, Carl P, Harper S, Dalhaimer P, Speicher DW, Discher DE (2003) Cooperativity in forced unfolding of tandem spectrin repeats. Biophys J 84:533–544

Le Rumeur E, Fichou Y, Pottier S, Gaboriau F, Rondeau-Mouro C, Vincent M, Gallay J, Bondon A (2003) Interaction of dystrophin rod domain with membrane phospholipids: evidence of a close proximity between tryptophan residues and lipids. J Biol Chem 278:5993–6001

Le Rumeur E, Pottier S, DA Costa G, Metzinger L, Mouret L, Rocher C, Fourage M, Rondeau-Mouro C, Bondon A (2007) Binding of the dystrophin second repeat to membrane di-oleyl phospholipids is dependent upon lipid packing. Biochim Biophys Acta 1768:648–654

Le Rumeur E, Winder SJ, Hubert JF (2010) Dystrophin: more than just the sum of its parts. Biochim Biophys Act 1804:1713–1722

Lecomte MC, Dhermy D, Garbarz M, Feo C, Gautero H, Bournier O, Picat C, Chaveroche I, Galand C, Boivin P (1987) Hereditary pyropoikilocytosis and elliptocytosis in a Caucasian family. Transmission of the same molecular defect in spectrin through three generations with different clinical expression. Hum Genet 77:329–334

Lecomte MC, Garbarz M, Gautero H, Bournier O, Galand C, Boivin P, Dhermy D (1993) Molecular basis of clinical and morphological heterogeneity in hereditary elliptocytosis (HE) with spectrin alpha I variants. Br J Haematol 85:584–595

Legardinier S, Hubert J-F, Le Bihan O, Tascon C, Rocher C, Raguénès-Nicol C, Bondon A, Hardy S, Le Rumeur E (2008) Sub-domains of the dystrophin rod domain display contrasting lipid-binding and stability properties. Biochim Biophys Acta 1784:672–682

Legardinier S, Raguénès-Nicol C, Tascon C, Rocher C, Hardy S, Hubert JF, Le Rumeur E (2009) Mapping of the lipid-binding and stability properties of the central rod domain of human dystrophin. J Mol Biol 389:546–558

Legrand B, Giudice E, Nicolas A, Delalande O, Le Rumeur E (2011) Computational study of the human dystrophin repeats: interaction properties and molecular dynamics. PLoS One 6:e23819

Leluk J, Hanus-Lorenz B, Sikorski AF (2001) Application of genetic semihomology algorithm to theoretical studies on various protein families. Acta Biochim Pol 48:21–33

Lemmon MA, Ferguson KM, Abrams CS (2002) Pleckstrin homology domains and the cytoskeleton. FEBS Lett 513:71–76

Lidov HG, Selig S, Kunkel LM (1995) Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. Hum Mol Genet 4:329–335

Lin AY, Prochniewicz E, James ZM, Svensson B, Thomas DD (2011) Large-scale opening of utrophin's tandem calponin homology (CH) domains upon actin binding by an induced-fit mechanism. Proc Natl Acad Sci U S A 108:12729–12733

Lorenzo DN, Li MG, Mische SE, Armbrust KR, Ranum LP, Hays TS (2010) Spectrin mutations that cause spinocerebellar ataxia type 5 impair axonal transport and induce neurodegeneration in Drosophila. J Cell Biol 189:143–158

Machnicka B, Czogalla A, Hryniewicz-Jankowska A, Boguslawska DM, Grochowalska R, Heger E, Sikorski AF (2014) Spectrins: a structural platform for stabilization and activation of membrane channels, receptors and transporters. Biochim Biophys Acta 1838:620–634

Macias MJ, Musacchio A, Ponstingl H, Nilges M, Saraste M, Oschkinat H (1994) Structure of the pleckstrin homology domain from beta-spectrin. Nature 369:675–677

Mcgreevy JW, Hakim CH, Mcintosh MA, Duan D (2015) Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech 8:195–213

Mehboob S, Song Y, Witek M, Long F, Santarsiero BD, Johnson ME, Fung LW (2010) Crystal structure of the nonerythroid alpha-spectrin tetramerization site reveals differences between erythroid and nonerythroid spectrin tetramer formation. J Biol Chem 285:14572–14584

Mirijanian DT, Chu JW, Ayton GS, Voth GA (2007) Atomistic and coarse-grained analysis of double spectrin repeat units: the molecular origins of flexibility. J Mol Biol 365:523–534

Molza A-E, Férey N, Czjzek M, Le Rumeur E, Hubert J-F, Tek A, Laurent B, Baaden M, Delalande O (2014) Innovative interactive flexible docking method for multi-scale reconstruction elucidates dystrophin molecular assembly. Faraday Discuss 169:45–62

Molza AE, Mangat K, Le Rumeur E, Hubert JF, Menhart N, Delalande O (2015) Structural basis of neuronal nitric-oxide synthase interaction with dystrophin repeats 16 and 17. J Biol Chem 290:29531–29541

Monaco A, Bertelson C, Liechti-Gallati S, Moser H, Kunkel L (1988) An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics 2:90–95

Musacchio A, Noble M, Pauptit R, Wierenga R, Saraste M (1992) Crystal structure of a Src-homology 3 (SH3) domain. Nature 359:851–855

Muthu M, Richardson KA, Sutherland-Smith AJ (2012) The crystal structures of dystrophin and utrophin spectrin repeats: implications for domain boundaries. PLoS One 7:e40066

Nans A, Mohandas N, Stokes DL (2011) Native ultrastructure of the red cell cytoskeleton by cryo-electron tomography. Biophys J 101:2341–2350

Nicolas G, Pedroni S, Fournier C, Gautero H, Craescu C, Dhermy D, Lecomte MC (1998) Spectrin self-association site: characterization and study of beta-spectrin mutations associated with hereditary elliptocytosis. Biochem J 332(Pt 1):81–89

Nicolas A, Lucchetti-Miganeh C, Ben Yaou R, Kaplan JC, Chelly J, Leturcq F, Barloy-Hubler F, Le Rumeur E (2012) Assessment of the structural and functional impact of in-frame mutations of the DMD gene, using the tools included in the eDystrophin online database. Orphanet J Rare Dis 7:45

Nicolas A, Delalande O, Hubert JF, Le Rumeur E (2014) The spectrin family of proteins: a unique coiled-coil fold for various molecular surface properties. J Struct Biol 186:392–401

Nicolas A, Raguenes-Nicol C, Ben Yaou R, Ameziane-Le Hir S, Cheron A, Vie V, Claustres M, Leturcq F, Delalande O, Hubert JF, Tuffery-Giraud S, Giudice E, Le Rumeur E (2015) Becker muscular dystrophy severity is linked to the structure of dystrophin. Hum Mol Genet 24:1267–1279

Nilges M, Macias MJ, O'donoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. J Mol Biol 269:408–422

Norwood F, Sutherland-Smith A, Keep N, Kendrick-Jones J (2000) The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy. Structure 8:481–491

Nudel U, Zuk D, Einat P, Zeelon E, Levy Z, Neuman S, Yaffe D (1989) Duchenne muscular dystrophy gene product is not identical in muscle and brain. Nature 337:76–78

Ogawa Y, Schafer DP, Horresh I, Bar V, Hales K, Yang Y, Susuki K, Peles E, Stankewich MC, Rasband MN (2006) Spectrins and ankyrinB constitute a specialized paranodal cytoskeleton. J Neurosci 26:5230–5239

Ohara O, Ohara R, Yamakawa H, Nakajima D, Nakayama M (1998) Characterization of a new beta-spectrin gene which is predominantly expressed in brain. Brain Res Mol Brain Res 57:181–192

Parry DAD, Dixon TW, Cohen C (1992) Analysis of the three-alpha-helix motif in the spectrin superfamily of proteins. Biophys J 61:858–867

Pascual J, Pfuhl M, Walther D, Saraste M, Nilges M (1997) Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. J Mol Biol 273:740–751

Petrof BJ, Shrager JB, Stedmann HH, Kelly AM, Sweeney HL (1993) Dystrophin protects the sarcolemma from stresses developed during muscle contraction. Proceedings of the national academy of sciences, USA, 90, 3710–3714

Prins KW, Humston JL, Mehta A, Tate V, Ralston E, Ervasti JM (2009) Dystrophin is a microtubule-associated protein. J Cell Biol 186:363–369

Rahimov F, Kunkel LM (2013) The cell biology of disease: cellular and molecular mechanisms underlying muscular dystrophy. J Cell Biol 201:499–510

Ribiero Ede A Jr, Pinotsis N, Ghisleni A, Salmazo A, Konarev PV, Kostan J, Sjöblum B, Schreiner C, Polyansky AA, Gkougkoulia EA, Holt MR, Aachnann FL, Zagrovic B, Bordignon E, Pirker KF, Svergun DI, Gautel M, Djinovic-Carugo K. (2014) The structure and regulation of human muscle α-actinin. Cell 159:1447–1460

Robertsson J, Petzold K, Löfvenberg L, Backman L (2005) Folding of spectrin's SH3 domain in the presence of spectrin repeats. Cell Mol Biol Lett 10(4):595–612, PubMed PMID: 16341269

Sahr KE, Laurila P, Kotula L, Scarpa AL, Coupal E, Leto TL, Linnenbach AJ, Winkelmann JC, Speicher DW, VT M et al (1990) The complete cDNA and polypeptide sequences of human erythroid alpha-spectrin. J Biol Chem 265:4434–4443

Saitsu H, Tohyama J, Kumada T, Egawa K, Hamada K, Okada I, Mizuguchi T, Osaka H, Miyata R, Furukawa T, Haginoya K, Hoshino H, Goto T, Hachiya Y, Yamagata T, Saitoh S, Nagai T, Nishiyama K, Nishimura A, Miyake N, Komada M, Hayashi K, Hirai S, Ogata K, Kato M, Fukuda A, Matsumoto N (2010) Dominant-negative mutations in alpha-II spectrin cause west syndrome with severe cerebral hypomyelination, spastic quadriplegia, and developmental delay. Am J Hum Genet 86:881–891

Sander M, Chavoshan B, Harris SA, Iannaccone ST, Stull JT, Thomas GD, Victor RG (2000) Functional muscle ischemia in neuronal nitric oxide synthase-deficient skeletal muscle of children with Duchenne muscular dystrophy. Proc Natl Acad Sci U S A 97:13818–13823

Sarkis J, Hubert JF, Legrand B, Robert E, Cheron A, Jardin J, Hitti E, Le Rumeur E, Vie V (2011) Spectrin-like repeats 11–15 of human dystrophin show adaptations to a lipidic environment. J Biol Chem 286:30481–30491

Sato K, Yokota T, Ichioka S, Shibata M, Takeda S (2008) Vasodilation of intramuscular arterioles under shear stress in dystrophin-deficient skeletal muscle is impaired through decreased nNOS expression. Act Myol 27:30–36

Seto JT, Ramos JN, Muir L, Chamberlain JS, Odom GL (2012) Gene replacement therapies for duchenne muscular dystrophy using adeno-associated viral vectors. Curr Gene Ther 12:139–151

Sikorski AF, Terlecki G, Zagon IS, Goodman SR (1991) Synapsin I-mediated interaction of brain spectrin with synaptic vesicles. J Cell Biol 114:313–318

Sikorski AF, Sangerman J, Goodman SR, Critz SD (2000) Spectrin (betaSpIIsigma1) is an essential component of synaptic transmission. Brain Res 852:161–166

Singh SM, Mallela KM (2012) The N-terminal actin-binding tandem calponin-homology (CH) domain of dystrophin is in a closed conformation in solution and when bound to F-actin. Biophys J 103:1970–1978

Stabach PR, Morrow JS (2000) Identification and characterization of beta V spectrin, a mammalian ortholog of Drosophila beta H spectrin. J Biol Chem 275:21385–21395

Stabach PR, Simonovic I, Ranieri MA, Aboodi MS, Steitz TA, Simonovic M, Morrow JS (2009) The structure of the ankyrin-binding site of beta-spectrin reveals how tandem spectrin-repeats generate unique ligand-binding properties. Blood 113:5377–5384

Stankewich MC, Tse WT, Peters LL, Ch'ng Y, John KM, Stabach PR, Devarajan P, Morrow JS, Lux SE (1998) A widely expressed betaIII spectrin associated with Golgi and cytoplasmic vesicles. Proc Natl Acad Sci U S A 95:14158–14163

Sutherland-Smith AJ, Moores CA, Norwood FL, Hatch V, Craig R, Kendrick-Jones J, Lehman W (2003) An atomic model for actin binding by the CH domains and spectrin-repeat modules of utrophin and dystrophin. J Mol Biol 329:15–33

Tinsley J, Blake D, Roche A, Fairbrother U, Riss J, Byth B, Knight A, Kendrick-Jones J, Suthers G, Love D, Edwards Y, Davies K (1992) Primary stucture of dystrophin- related protein. Nature 360:591–593

Tuffery-Giraud S, Beroud C, Leturcq F, Yaou RB, Hamroun D, Michel-Calemard L, Moizard MP, Bernard R, Cossee M, Boisseau P, Blayau M, Creveaux I, Guiochon-Mantel A, De Martinville B, Philippe C, Monnier N, Bieth E, Khau Van Kien P, Desmet FO, Humbertclaude V, Kaplan JC, Chelly J, Claustres M (2009) Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum Mutat 30:934–945

Voas MG, Lyons DA, Naylor SG, Arana N, Rasband MN, Talbot WS (2007) alphaII-spectrin is essential for assembly of the nodes of Ranvier in myelinated axons. Curr Biol 17:562–568

Waller KL, Nunomura W, An X, Cooke BM, Mohandas N, Coppel RL (2003) Mature parasite-infected erythrocyte surface antigen (MESA) of Plasmodium falciparum binds to the 30-kDa domain of protein 4.1 in malaria-infected red blood cells. Blood 102:1911–1914

Wein N, Alfano L, Flanigan KM (2015) Genetics and emerging treatments for Duchenne and Becker muscular dystrophy. Pediatr Clin N Am 62(3):723–742

Wichterle H, Hanspal M, Palek J, Jarolim P (1996) Combination of two mutant alpha spectrin alleles underlies a severe spherocytic hemolytic anemia. J Clin Invest 98:2300–2307

Wilmotte R, Harper SL, Ursitti JA, Marechal J, Delaunay J, Speicher DW (1997) The exon 46-encoded sequence is essential for stability of human erythroid alpha-spectrin and heterodimer formation. Blood 90:4188–4196

Wilton SD, Veedu RN, Fletcher S (2015) The emperor's new dystrophin: finding sense in the noise. Trends Mol Med 21(7):417–426

Winder SJ, Gibson TJ, Kendrick-Jones J (1995) Dystrophin and utrophin: the missing links! FEBS Lett 369:27–33

Winkelmann JC, Chang JG, Tse WT, Scarpa AL, Marchesi VT, Forget BG (1990) Full-length sequence of the cDNA for human erythroid beta-spectrin. J Biol Chem 265:11827–11832

Wolny M, Grzybek M, Bok E, Chorzalska A, Lenoir M, Czogalla A, Adamczyk K, Kolondra A, Diakowski W, Overduin M, Sikorski AF (2011) Key amino acid residues of ankyrin-sensitive phosphatidylethanolamine/phosphatidylcholine-lipid binding site of betaI-spectrin. PLoS One 6:e21538

Writzl K, Primec ZR, Strazisar BG, Osredkar D, Pecaric-Meglic N, Kranjc BS, Nishiyama K, Matsumoto N, Saitsu H (2012) Early onset West syndrome with severe hypomyelination and coloboma-like optic discs in a girl with SPTAN1 mutation. Epilepsia 53:e106–e110

Xu K, Zhong G, Zhuang X (2013) Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. Science 339:452–456

Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D (1993) Crystal structure of the repetitive segments of spectrin. Science 262:2027–2030

Zhang P, Talluri S, Deng H, Branton D, Wagner G (1995) Solution structure of the pleckstrin homology domain of drosophila beta-spectrin. Structure 3:1185–1195

Zhang Z, Weed SA, Gallagher PG, Morrow JS (2001) Dynamic molecular modeling of pathogenic mutations in the spectrin self-association domain. Blood 98:1645–1653

Zimmer WE, Zhao Y, Sikorski AF, Critz SD, Sangerman J, Elferink LA, Xu XS, Goodman SR (2000) The domain of brain beta-spectrin responsible for synaptic vesicle association is essential for synaptic transmission. Brain Res 881:18–27

Chapter 13
Fibrin Formation, Structure and Properties

John W. Weisel and Rustem I. Litvinov

Contents

J.W. Weisel (✉) • R.I. Litvinov
Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
e-mail: weisel@mail.med.upenn.edu

Abstract   Fibrinogen and fibrin are essential for hemostasis and are major factors in thrombosis, wound healing, and several other biological functions and pathological conditions. The X-ray crystallographic structure of major parts of fibrin(ogen), together with computational reconstructions of missing portions and numerous biochemical and biophysical studies, have provided a wealth of data to interpret molecular mechanisms of fibrin formation, its organization, and properties. On cleavage of fibrinopeptides by thrombin, fibrinogen is converted to fibrin monomers, which interact via knobs exposed by fibrinopeptide removal in the central region, with holes always exposed at the ends of the molecules. The resulting half-staggered, double-stranded oligomers lengthen into protofibrils, which aggregate laterally to make fibers, which then branch to yield a three-dimensional network. Much is now known about the structural origins of clot mechanical properties, including changes in fiber orientation, stretching and buckling, and forced unfolding of molecular domains. Studies of congenital fibrinogen variants and post-translational modifications have increased our understanding of the structure and functions of fibrin(ogen). The fibrinolytic system, with the zymogen plasminogen binding to fibrin together with tissue-type plasminogen activator to promote activation to the active proteolytic enzyme, plasmin, results in digestion of fibrin at specific lysine residues. In spite of a great increase in our knowledge of all these interconnected processes, much about the molecular mechanisms of the biological functions of fibrin(ogen) remains unknown, including some basic aspects of clotting, fibrinolysis, and molecular origins of fibrin mechanical properties. Even less is known concerning more complex (patho)physiological implications of fibrinogen and fibrin.

Keywords   Fibrin formation • Fibrin structure • Fibrin properties • Fibrinogen composition • α-Helical coiled-coil • Blood clot • Fibrin polymerization • Clot mechanical properties • Molecular mechanisms of fibrinolysis • Modulation of clot structure

Abbreviations

FpA and FpB   fibrinopeptides A and B
GHRP   Gly-His-Arg-Pro
GPRP   Gly-Pro-Arg-Pro
NXS or NXT   Asn-X-Ser or Asn-X-Thr

Plg   plasminogen
Pn   plasmin
t-PA   tissue-type Plg activator

13.1   Introduction

Fibrinogen was first classified as a fibrous protein with keratin, myosin and epidermin, based on its wide angle X-ray diffraction pattern arising from its α-helical coiled-coil structure (Bailey et al. 1943). It is a 340-kDa glycoprotein, normally present in human blood plasma at a concentration of about 1.5–4 g/L, that is essential for hemostasis, wound healing, inflammation, angiogenesis, and several other biological functions. Fibrinogen is a soluble macromolecule, but forms an insoluble clot or gel on conversion to fibrin by the action of the serine protease thrombin, which is activated by a cascade of enzymatic reactions triggered by vessel wall injury, activated blood cells, or a foreign surface (Fig. 13.1). A mechanically stable clot is necessary to prevent blood loss (stopping bleeding is called hemostasis) and to promote wound healing. Fibrin clots are dissolved by the fibrinolytic system, acting in a series of enzymatic reactions with positive and negative feedback.

   In vivo, there is a careful balance between clotting, the conversion of fibrinogen to fibrin, and fibrinolysis, the proteolytic dissolution of the clot (Fig. 13.1). Imbalance in one direction (prevalence of fibrinolysis) can lead to bleeding while the opposite imbalance (prevalence of clotting) can cause thrombosis, or formation of a clot that blocks the flow of blood through a vessel (called a thrombus). Thrombosis, often resulting from atherosclerosis or many other pathological pro-
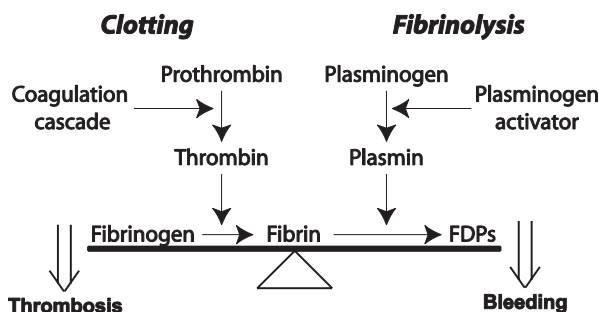
Fig. 13.1 Basic scheme of fibrin clot formation and fibrinolysis and the balance between these processes. The clot is formed via a cascade of enzymatic reactions that activates prothrombin to the proteolytic enzyme thrombin, which converts soluble fibrinogen to make insoluble fibrin, the process referred to as blood clotting. The fibrin clot is dissolved through fibrinolysis or cleavage by the proteolytic enzyme plasmin, resulting in fibrin degradation products (FDPs). Plasmin is formed on the fibrin surface from the zymogen plasminogen by plasminogen activators. There is a balance between clotting and fibrinolysis such that excess clotting can lead to thrombosis, while excess fibrinolysis can lead to bleeding

cesses, is the most common cause of myocardial infarction, ischemic stroke, deep vein thrombosis, and other cardiovascular diseases.

In addition to fibrin clot formation, fibrinogen is also necessary for an earlier step in hemostasis (called "primary hemostasis"), the aggregation of platelets leading to formation of a platelet "plug" at the site of vessel wall injury. The bivalent fibrinogen molecules act as bridges to link activated platelets, since the ends of rod-like fibrinogen bind with high affinity to the major adhesive receptor on platelets, the integrin αIIbβ3. Fibrinogen also binds specifically to some other cells, but none of these cellular interactions will be discussed here, though other reviews on this topic are available (Bennett 2001; Wei et al. 2009; Coller and Shattil 2008; Coller 2011).

Fibrinogen and fibrin were last reviewed in the previous series of books dedicated entirely to Fibrous Proteins in 2005 (Weisel 2005), but many other reviews have appeared during the past 10 years that have summarized various aspects of the biology and biochemistry of fibrinogen and fibrin, though none as broad in scope (Cilia La Corte et al. 2011; Wolberg 2010, 2012; Undas and Ariens 2011; Ariens 2013; Weisel and Litvinov 2013; Weisel and Dempfle 2013; Bridge et al. 2014; Lord 2007, 2011; Falvo et al. 2010). We now know more about the process of fibrin formation, modulation of clot properties, and fibrinolysis, but perhaps the greatest explosion of new knowledge has been related to the structural origin of clot mechanical properties. Of special note for fibrous proteins, old observations of the α-helix to β-sheet transition upon stretching of fibrin have been confirmed and a functional spring-like role for the α-helical coiled-coil structure has been discovered. Although most data referenced in this review are related to human fibrinogen and fibrin, the basic principles discussed here have general relevance and are beyond species-specific peculiarities.

13.2 Biochemistry of Fibrinogen, the Precursor to Fibrin

13.2.1 Biosynthesis of Fibrinogen in Hepatocytes

Fibrinogen is the product of three closely linked genes, *FGA*, *FGB*, and *FGG*, each specifying the primary structure of one of its three polypeptide chains, Aα, Bβ, and γ, respectively (Chung et al. 1983, 1990; Crabtree 1987). The fibrinogen genes are clustered on human chromosome 4 (Kant et al. 1985) and they translate into nascent polypeptides of pre-pro-Aα chain (644 amino acid residues), pre-pro-Bβ chain (491 residues), and pre-pro-γ chain (437 residues). An extension of an open reading frame into an alternately spliced sixth intron gives rise to a longer Aα chain (αE) in 1–2 % of molecules in the adult (more in fetal blood), so that they contain an additional domain that is homologous to the C-terminal Bβ and γ chains, resulting in a fibrinogen molecule with a molecular mass of 420 kDa compared to 340 kDa for the major fibrinogen fraction with the shorter Aα chains (Fu and Grieninger 1994). The alternative splicing of a γ chain mRNA produces 8–15 % of plasma fibrinogen

molecules in which the γ chain C-terminal 400–411 dodecapeptide (γA chain) is altered by adding new amino acids from 408 to 427 (γ′ chain) (Wolfenstein and Mosesson 1981; Chung and Davie 1984; de Maat and Verschuur 2005).

Before fibrinogen is secreted from hepatocytes into blood, it needs to undergo several steps of assembly of the polypeptide chains. Following translation of each of the chains, independent translocation into the lumen of the endoplasmic reticulum, interactions of the chains with chaperone proteins that assist in the assembly and folding processes, and quality control mechanisms that distinguish properly assembled fibrinogen destined for secretion from unassembled forms that are degraded intracellularly (Redman and Xia 2001; Tamura et al. 2013). During translocation of the single polypeptides into the lumen of the endoplasmic reticulum, a signal peptide is cleaved from each chain. After full processing and assembly in the endoplasmic reticulum and Golgi organelles, there are 610, 461, and 411 amino acids in each of the common final forms of the human Aα, Bβ and γ chains, respectively, remaining from the corresponding nascent pre-pro-polypeptides. In the endoplasmic reticulum there is a progression from single chains to two-chain complexes to trimeric half molecules generated by combining a Bβ chain with the Aα-γ dimer or an Aα chain with the Bβ-γ dimer. Next, the two trimers are joined at their N-termini to form the hexamer via disulfide bridges necessary for assembly of the two half molecules (Zhang and Redman 1996). Although the distal clusters of interchain disulfide bonds (named "disulfide rings") are not necessary for assembly of the two half molecules, they are necessary for secretion. In addition to assembly and formation of disulfide bonds, before secretion fibrinogen undergoes a number of co-translational and post-translational modifications, including N-glycosylation, these are discussed below. The structure of fibrinogen can be modified even after it has been secreted into the blood, for example by limited proteolysis and glycation.

13.2.2 Fibrinogen Metabolism

The liver is the primary source of plasma fibrinogen, with a steady state rate of synthesis of 1.7–5 g per day and a large intracellular reserve (Takeda 1966). Three quarters of human fibrinogen is present in the plasma but it is also in platelets, lymph and interstitial fluid. The normal half-life of fibrinogen is 3–5 days, but despite the numerous studies on the distribution of iodine-labeled fibrinogen, its physiological catabolic pathway is largely unknown. Coagulation and fibrinolysis accounts for only 2–3 % of fibrinogen loss in healthy individuals (Nossel 1976). Moreover, fibrinogen turnover in the presence and absence of heparin, a potent inhibitor of blood clotting, did not reveal any differences in the half-life of fibrinogen, so intravascular fibrin formation was not found to be a quantitatively significant route in fibrinogen catabolism under physiological conditions in humans (Takeda 1966; Collen et al. 1972). Another proposed pathway of fibrinogen catabolism is proteolytic degradation by plasmin; however, in the presence of tranexamic acid, an inhibitor of fibrinolysis, the half-life of labeled fibrinogen did not change (Collen

et al. 1972). Nevertheless, proteolytic fibrin(ogen) degradation products appear to play a role in the regulation of fibrinogen turnover (Nham and Fuller 1986).

Fibrinogen is one of the acute phase proteins, which are up-regulated in response to injury and inflammation, followed by an up to ten-fold increase in its concentration in the blood (Crabtree 1987). The up-regulation is mediated largely by interleukin-6 and perhaps other pro-inflammatory mediators that trigger intracellular signaling pathways in hepatocytes and modulate gene expression via various transcription factors (Hantgan et al. 2000; Fish and Neerman-Arbez 2012).

There is intracellular fibrinogen stored in platelet α-granules, but its origin is controversial, as is whether it is structurally and functionally distinct from plasma fibrinogen. A separate origin is suggested by the observation that the γ chain variants of fibrinogen appear to be absent in platelets (Haidaris et al. 1989), and in one patient with a heterozygous genetically modified fibrinogen, the platelets contained only normal fibrinogen (Jandrot-Perrus et al. 1979). Since platelets originate from megakaryocytes, the latter are also a possible source of fibrinogen synthesis. However, in subjects with a low fibrinogen level in blood (hypofibrinogenemia), there are also lower levels of fibrinogen in platelets, and infusion of fibrinogen results in a subsequent increase in platelet fibrinogen (Harrison et al. 1989). Also, no fibrinogen mRNA has been found in megakaryocytes (Louache et al. 1991). Therefore, it appears that platelet and megakaryocyte fibrinogen arise primarily from αIIbβ3-mediated internalization of plasma fibrinogen (Harrison et al. 1990).

Although the liver is the primary source of plasma fibrinogen, fibrinogen is also synthesized in some extra-hepatic tissues. Only fibrinogen γ chain gene expression has been shown *in vivo* in bone marrow, brain, and lung (Haidaris and Courtney 1990). Epithelial cells from lung and intestine secrete small amounts of fibrinogen in a polarized manner from their basolateral face (Haidaris 1997). It is possible that lung epithelium secretes fibrinogen and incorporates it into the extracellular matrix under certain pathological conditions, contributing to fibrotic lung disease, because the amount of fibrinogen expressed in lung epithelial cells is dramatically increased after treatments with dexamethasone and interleukin-6 (Lawrence and Simpson-Haidaris 2004). Synthesis of fibrinogen by cultured granulosa cells may reflect a possible function for it in ovulation (Parrott et al. 1993). The apparent *in vivo* synthesis of fibrinogen by trophoblasts (Galanakis et al. 1996) and the fact that the trophoblast basement membrane consists largely of fibrin(ogen) suggest that these cells may secrete fibrinogen into their abluminal and/or interstitial environment, but the functional significance is as yet unknown. Taken together, the normal biological relevance of the synthesis of fibrinogen in extra-hepatic tissues is unclear, but it may become important under certain pathological circumstances.

13.2.3 *Polypeptide Chain Composition of Fibrin(ogen)*

Human fibrinogen is made up of three pairs of polypeptide chains, designated Aα, Bβ and γ, with molecular masses of 66,500, 52,000, and 46,500 Da, respectively (Fig. 13.2). The co- and post-translational addition of N-linked carbohydrate to the
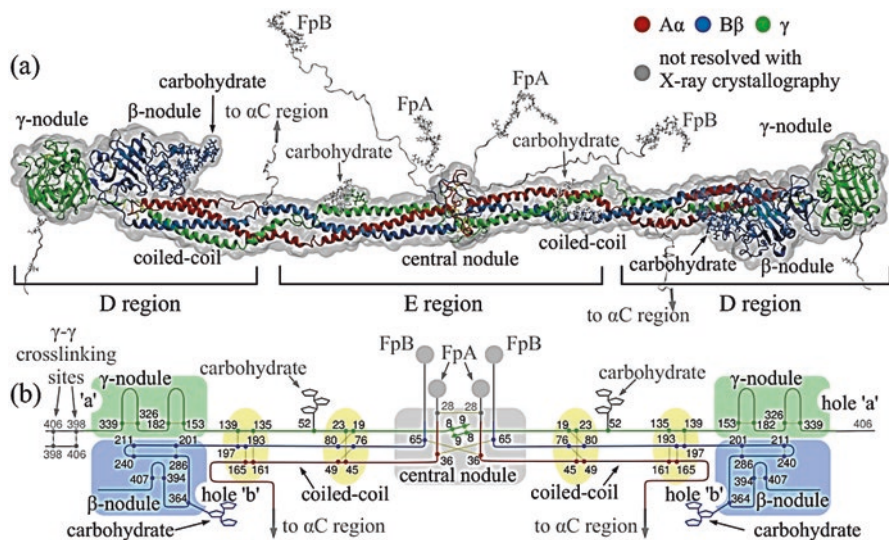
Fig. 13.2 Fibrinogen structure. (**a**) The atomic resolution structure of about two-thirds of the fibrinogen molecule has been determined by X-ray crystallography (PDB Entry: 3GHG). Fibrinogen and its parts are shown with addition of portions missing from the crystal structure reconstructed computationally, namely the amino terminal ends of the Aα and Bβ chains with FpA and FpB in the central nodule and the beginning of the αC regions. (**b**) Schematic diagram of the polypeptide chains of fibrinogen. The Aα, Bβ and γ chains are represented by lines with lengths proportional to the number of amino acid residues in each chain and various structural regions are labeled (Zhmurov et al. 2011, with permission from Elsevier Ltd.)

Bβ and γ chains brings the total molecular mass to about 340 kDa. The nomenclature for the polypeptide composition of fibrinogen (Aα Bβ γ)₂, arises from the designation of the small fibrinopeptides A and B (FpA and FpB) that comprise the N-terminal ends of the Aα and Bβ chains, respectively, and are cleaved by thrombin to yield the α and β chains without the fibrinopeptides. No peptides are cleaved by thrombin from the γ chains, hence the subunit composition of monomeric fibrin is (α β γ)₂ and the conversion of fibrinogen into fibrin monomer can be described as (Aα Bβ γ)₂ → (α β γ)₂ + 2FpA + 2FpB.

All six chains are held together by 29 disulfide bonds (Henschen and McDonagh 1986) to make two symmetrical half-molecules (Fig. 13.2). There are 8, 11, and 10 cysteine residues in the Aα, Bβ, and γ chains, respectively, and the amino termini of all six chains are held together by disulfide bonds in the central globule. Unusual Cys-Pro-X-X-Cys sequences occurring twice in each chain are arranged into "disulfide ring" structures, in which all three chains are joined together at each end of the α-helical coiled-coils (Doolittle 1984). Three interchain disulfide bonds link the two halves of the molecule together, one between the two Aα chains and two between the two γ chains. A single interchain disulfide bond connects the Aα and Bβ chains within each half-molecule. The remainder of the Aα chain contains one intrachain disulfide, while the Bβ chain contains three intrachain disulfides and the γ chain contains two.

13.2.4   Overall Structure of Fibrinogen Molecules

Based on transmission electron microscopy, atomic force microscopy, and X-ray crystallographic data, the fibrinogen molecule has an elongated shape 45 nm in length and ~2–5 nm in diameter (Fig. 13.2; Hall and Slayter 1959; Fowler and Erickson 1979; Williams 1981; Weisel et al. 1985). Plasmin cleaves fibrinogen into a number of core fragments, namely fragment E comprising the central part of the molecule and two identical fragments D originating from the lateral parts of fibrinogen (Nussenzweig et al. 1961). According to the nomenclature of the corresponding proteolytic fragments, the fibrinogen molecule has two distal globular D regions and one central globular E region each containing a part of α-helical coiled-coils (Medved and Weisel 2009) (Fig. 13.2). The atomic resolution structure of more than two-thirds of fibrinogen has been accomplished through X-ray crystallography (Yee et al. 1997; Spraggon et al. 1997; Brown et al. 2000; Madrazo et al. 2001), though the structure still has not been completed, since there are missing unstructured portions, which are not resolved crystallographically. These are residues 1–26, 1–57, and 1–13 at the N-termini of the Aα, Bβ and γ chains, respectively, and residues 201–610, 459–461, and 395–411 at the C-termini of the Aα, Bβ and γ chains, respectively. These unresolved portions have been reconstructed computationally (Zhmurov et al. 2011) to gain a complete molecular structure of fibrinogen. Unlike other crystallographically unresolved parts of fibrinogen, the 390 residue-long C-termini of the Aα chains (residues 221–610), named the αC regions, were visualized using transmission electron microscopy (Veklich et al. 1993) and atomic force microscopy (Protopopova et al. 2015). The αC regions fold back from the distal ends of the triple coiled-coils to form a fourth strand and then extend outward via a flexible connector to relatively compact C-terminal domains that interact with the central globule of fibrinogen (Veklich et al. 1993; Litvinov et al. 2007b). Truncation of the αC regions affects substantially the hydrodynamic behavior of fibrinogen (Raynal et al. 2013).

13.2.5   Domain Structure of Fibrinogen

Fibrinogen is organized into domains, or independently folded structural units (Weisel 2005). From the X-ray crystallographic studies of fibrinogen, the central E region has four domains and each D region comprises seven domains (Medved and Weisel 2009; Fig. 13.2). The E region is composed of two symmetrical parts, in which the C-terminal parts of the Aα, Bβ and γ chains form a *coiled-coil-E domain* comprising a triple α-helical structure. The N-terminal parts of two γ chains form an asymmetric *γN-domain* domain in the center of the E region. Another domain, called the *funnel-shaped domain*, is formed on the opposite side of the E region by the parts of two Aα and two Bβ chains. The globular part of the E region without the

coiled-coil-E domains is often called the *central nodule*. In the lateral D region, the N-terminal parts of all the Aα, Bβ and γ chains form a *coiled-coil-D domain* comprising a triple α-helical structure. The C-terminal parts of the β and γ chains make up the *β-nodule* and *γ-nodule*, respectively. They are also named β- or γ-modules and each is made of three domains. As determined in the crystal structure of the γ-module these domains are named *A-domain (*N-terminal), *B-domain* (central) and *P-domain* (C-terminal; Medved and Weisel 2009).

Although the structure of the remaining fibrinogen regions not present in the crystal structures is not as well defined, it is known from NMR studies that in human fibrinogen each αC region (residues Aα221–610) forms two relatively compact structures. Namely, the C-terminal part (residues Aα392–610) contains a distinct structure connected to the rest of the molecule via the N-terminal portion (residues Aα221–391) comprising a flexible unstructured tether. As a result, the compact part is referred to as the *αC-domain* and the flexible part as the *αC-connector*. Notably, the two unstructured *BβN regions* comprising the flexible N-terminal portions of the Bβ chains contain a number of functionally important binding sites (Gorkun et al. 2006).

13.2.6   *α-Helical Coiled-Coils of Fibrinogen*

The central nodule and end globular parts of fibrin(ogen) are joined together by 17-nm-long triple (and partially quadruple) α-helical coiled-coil connectors formed by 111 or 112 amino acid residues from each of the Aα Bβ and γ chains that are bordered by "disulfide rings" (Fig. 13.2). In the α-helical coiled-coil, three right-handed α-helices wind around each other to form a left-handed supercoil (Cohen and Parry 1990). Unlike most triple helices, each of the coiled-coils of fibrinogen has a fourth helix in the bundle containing 30 residues (AαSer166 to AαPro195) that begins at the lateral disulfide ring where the Aα chain makes a U-turn and stretches in the reverse direction for about ¼ of the length of the coiled-coil connector (Spraggon et al. 1997). The coiled-coils can bend around a central hinge point located in a non-helical segment of the γ chain adjacent to the carbohydrate attachment site γAsn52 (Marsh et al. 2013). The bent portion of the coiled-coil exposes the cleavage sites that can be hydrolyzed by plasmin and other proteases followed by degradation of fibrinogen into proteolytic fragments. It also explains the conformational flexibility of fibrinogen observed in solution and at interfaces (Kohler et al. 2015). The functional role of the coiled-coil region has been recently ascribed to the tensile deformation of fibrin fibers, namely to the ability to undergo partial unraveling and spring-like reversible extension-contraction that helps to accommodate and propagate the tensile stress along the fiber axis (Zhmurov et al. 2011). At a high extent of tensile or compressive deformation, the α-helices undergo mechanical conversion into β-sheets (Litvinov et al. 2012; Zhmurov et al. 2012; see Sect. 13.5.5).

13.2.7 Ca²⁺-Binding Sites in Fibrinogen

Fibrinogen has both strong and weak binding sites for calcium ions, which are important for its functions, including fibrin polymerization, and lytic stability. Although biochemical experiments originally suggested that there were three high-affinity calcium binding sites in fibrinogen (Marguerie et al. 1977), the binding sites were identified from X-ray crystallographic structures. High-affinity binding sites (named γ1) for calcium ions are present in the γ chains and are associated with four coordinating amino acid residues, namely γAsp318, γAsp320, γGly324, and γPhe322, and two strongly bound water molecules (Yee et al. 1997; Spraggon et al. 1997). Other high-affinity Ca²⁺-binding sites (named β1) are located in the β-nodules in loop β381–385, each of which has one coordinating water molecule (Everse et al. 1998b). The dissociation constant for calcium ions in these binding sites is high enough that both types of these sites will be fully occupied in fibrinogen at physiological Ca²⁺ concentrations.

Two other Ca²⁺-binding sites named γ2 and β2 have much lower affinities. The γ2 sites are located in the loops γ294–301 (Everse et al. 1999). Impairment of these sites by mutating residues γAsp298 and γAsp301 caused only moderate effects on the crystal structure and functional properties of fibrinogen (Kostelansky et al. 2007). It is likely that the γ2 sites are formed as a result of molecular rearrangements induced by crystal packing (Kostelansky et al. 2004a, b). The low-affinity Ca²⁺-binding sites β2 are formed by residues BβAsp261, BβAsp398, and γGlu132 and the backbone carbonyl oxygen of BβAsp263 (Everse et al. 1999; Kostelansky et al. 2002, 2004a, b). The β2 sites anchor the β-nodules to the coiled-coil connector and were shown to be involved in the lateral aggregation of protofibrils (Kostelansky et al. 2004b). It has been proposed that the β2 sites regulate accessibility of the tissue plasminogen activator (t-PA)-binding site in fibrin(ogen) (Doolittle and Pandi 2006). There may also be some low affinity Ca²⁺-binding sites associated with the sialic acid residues on the carbohydrate chains (Dang et al. 1989).

When calcium ions are bound to the γ chain high affinity sites, the γ chains are protected from enzymatic degradation (Ly and Godal 1973; Odrljin et al. 1996), similar to the way that the peptide Gly-Pro-Arg-Pro protects the molecule from plasmin digestion (Yamazumi and Doolittle 1992). Ca²⁺-binding does not affect FpA cleavage by thrombin or batroxobin, but appears to be important for modulating fibrin polymerization by enhancing lateral aggregation to form thicker fibers, so that mutations affecting the high-affinity Ca²⁺-binding site have severe functional consequences (Brennan et al. 2007). There is a conformational change in fibrin associated with FpB cleavage that is Ca²⁺-dependent (Dyr et al. 1989; Donovan and Mihalyi 1985; Mihalyi 1988). There is also a change in affinity of Ca²⁺-binding sites associated with FpB release and the accompanying conformational change. Gly-His-Arg-Pro, a knob 'B' mimetic peptide, binds 10-fold more strongly to fibrinogen in the presence of Ca²⁺ than in its absence, which may be related to the β2 site that is involved in the conformational change associated with binding of Gly-His-Arg-Pro (Everse et al. 1999).

13.2.8   Carbohydrate Moieties of Fibrinogen

Four oligosaccharide chains are linked to each molecule of fibrinogen by way of N-glycosidic bonds: two are connected to the BβAsn364 residues in the β-nodule (resolved in the crystal structure 3GHG) and the other two are connected to γAsn52 in the coiled-coils (unresolved crystallographically) (Fig. 13.2). These carbohydrate attachment sites contain the classic NXS or NXT sequences that are typical of N-glycosylation. On the other hand, the Aα chains are devoid of any carbohydrate, in spite of the presence of two NXS sequences. Variable desialylation, or removal of the terminal N-acetylneuraminic acid residue (sialyl), accounts for part of the heterogeneity of circulating fibrinogen. Fibrinogen isolated from human plasma contains equal amounts of mono- and di-sialylated carbohydrate chains but no asialo-chains (Townsend et al. 1982, 1984).

The carbohydrate on fibrinogen has striking consequences for fibrin polymerization and for clot structure. Patients with cirrhosis of the liver and some other liver diseases have fibrinogen with high levels of sialylation of their carbohydrate, resulting in fibrin networks containing thinner fibers with a higher density of branch points (Martinez et al. 1983). These results are consistent with studies using neuraminidase to remove sialic acid from the carbohydrate of normal fibrinogen, producing clots made up of thicker fibers (Dang et al. 1989). Fibrinogen synthesized in inflammatory conditions (acute phase fibrinogen) has substantially different oligosaccharide structure (Brennan 2015). Complete removal of carbohydrate has more striking effects on clot structure, resulting in clots made up of very thick fibers (Langer et al. 1988). These results suggest that both the charge and mass of the carbohydrate help to modulate the extent of lateral aggregation and that the carbohydrate moieties significantly enhance the solubility of fibrinogen. Based on computational reconstructions, it has been recently proposed that the bulky carbohydrate moieties can potentially affect the 'B-b' knob-hole interactions either by tethering the knob 'B' (γAsn52) or obstructing the hole 'b' (BβAsn364), or both (Zhmurov et al. in preparation).

13.3   Molecular Mechanisms of the Conversion of Fibrinogen to Fibrin

13.3.1   General Remarks

Conversion of fibrinogen to fibrin is one of the major consequences of the enzymatic cascade of blood coagulation that is essential for stopping bleeding (hemostasis), as well as in vascular obstruction or thrombosis. It occurs in two major steps: enzymatic and non-enzymatic (Fig.13.3). At the enzymatic step, there is thrombin-catalyzed cleavage of the fibrinopeptides of fibrinogen to form the fibrin monomer. Thrombin is a highly specific serine protease upon activation of its zymogen,
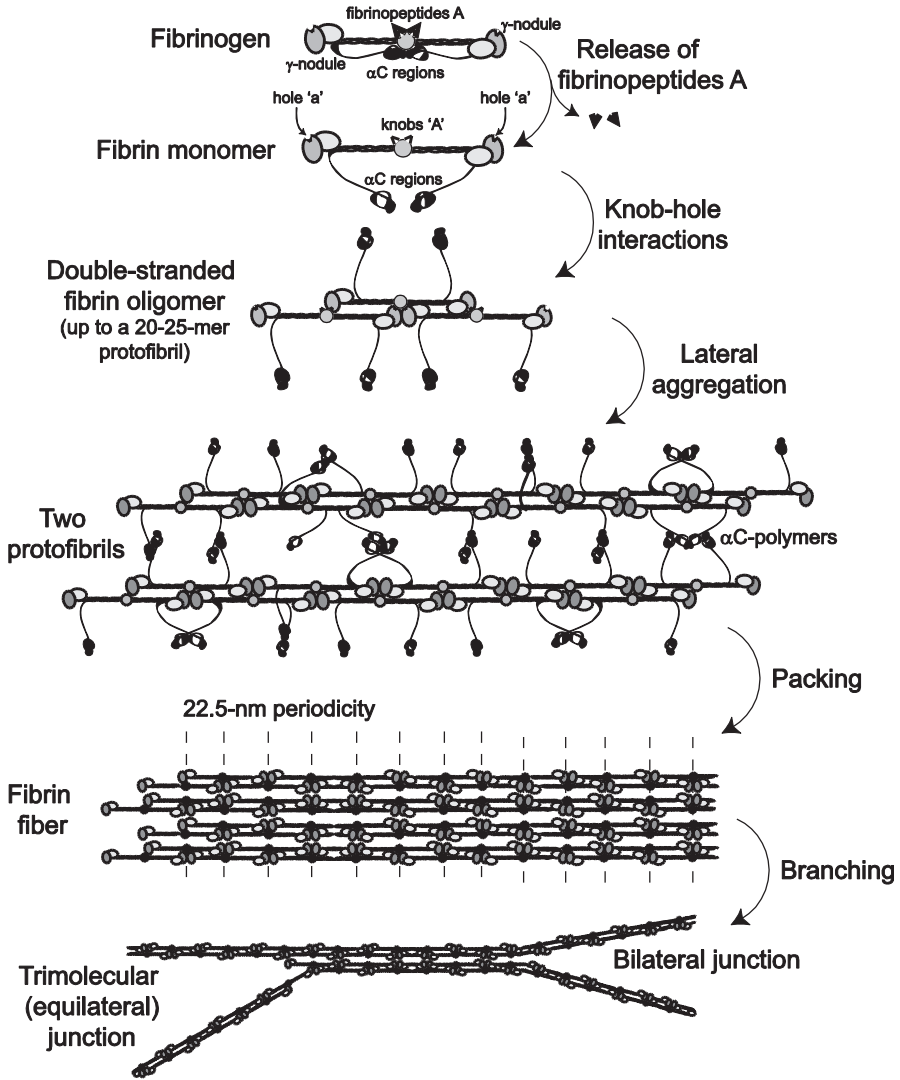
Fig. 13.3 Schematic diagram of fibrin polymerization. Fibrinopeptides in the central nodule cover knobs that are complementary to holes that are always exposed at the ends of the protein. When the fibrinopeptides are removed by thrombin, knob-hole interactions occur, giving rise to oligomers (a trimer 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 2013; Weisel and Dempfle 2013)

prothrombin, normally present in the blood. At the non-enzymatic stage, the monomeric fibrin self-assembles spontaneously to yield fibrin oligomers that lengthen to make two-stranded protofibrils. Protofibrils aggregate both laterally and longitudinally to form fibers that branch to yield a three-dimensional gelled network called a clot. Finally, the fibrin polymer is stabilized via covalent crosslinking by a plasma transglutaminase, Factor XIIIa, to form a mechanically and chemically more stable mature fibrin clot.

13.3.2   Enzymatic Release of Fibrinopeptides from Fibrinogen

Fibrin polymerization is triggered when thrombin cleaves FpA and FpB from the N-terminal portions of the Aα and Bβ chains of fibrinogen, respectively, producing monomeric fibrin. FpA (residues 1–16) is cleaved at the AαArg16-Gly17 peptide bond, while FpB (residues 1–14) is cleaved at the BβArg14-Gly15 bond, albeit more slowly than FpA. With polymerization, the rate of release of FpBs increases, reaching maximum when polymerization is almost complete, indicating that they are preferentially released from fibrin polymer (Erickson and Fowler 1983; Weisel et al. 1993). In surface-attached fibrinogen, unlike in solution, FpBs are cleaved at a faster rate than FpAs, depending on fibrinogen surface density and orientation (Riedel et al. 2011), indicating that the conformation of fibrinogen determines the ability of thrombin to access and cleave FpAs and FpBs differentially. This distinction in the cleavage rate of FpA and FpB is based on the spatial restrictions of the binding of thrombin to fibrinogen, implying that the N-termini of the Aα chain containing FpA are more accessible to the active site of thrombin (Pechik et al. 2006). When thrombin cleavage sites are mutated at positions AαArg16 or BβArg14 (in dysfibrinogenemias or recombinant fibrinogens) the release of FpA or FpB is precluded, leading to impaired fibrin formation (Galanakis et al. 1989; Moen et al. 2003).

13.3.3   'A-a' Knob-Hole Interactions in Fibrin

After cleavage of FpAs, the α chains have new N-terminal sequences Gly-Pro-Arg-(GPR) named knobs 'A' (Medved and Weisel 2009). During fibrin polymerization, knobs 'A' interact with holes or pockets 'a' that are always open in the γ-nodules of the interacting fibrin molecules; the binding of knobs 'A' to holes 'a' is called the 'A-a' interaction (Figs. 13.3 and 13.4). The 'A-a' interactions have been characterized biophysically at the single-molecule level as quite strong, highly specific, and stable intermolecular associations (Litvinov et al. 2005).

Structures of fragment D (corresponding to the lateral D regions containing the γ-nodule) co-crystallized with the peptide GPRP (synthetic knob 'A' mimetic) clearly indicated that this ligand was located in hollows present in the γ-nodule
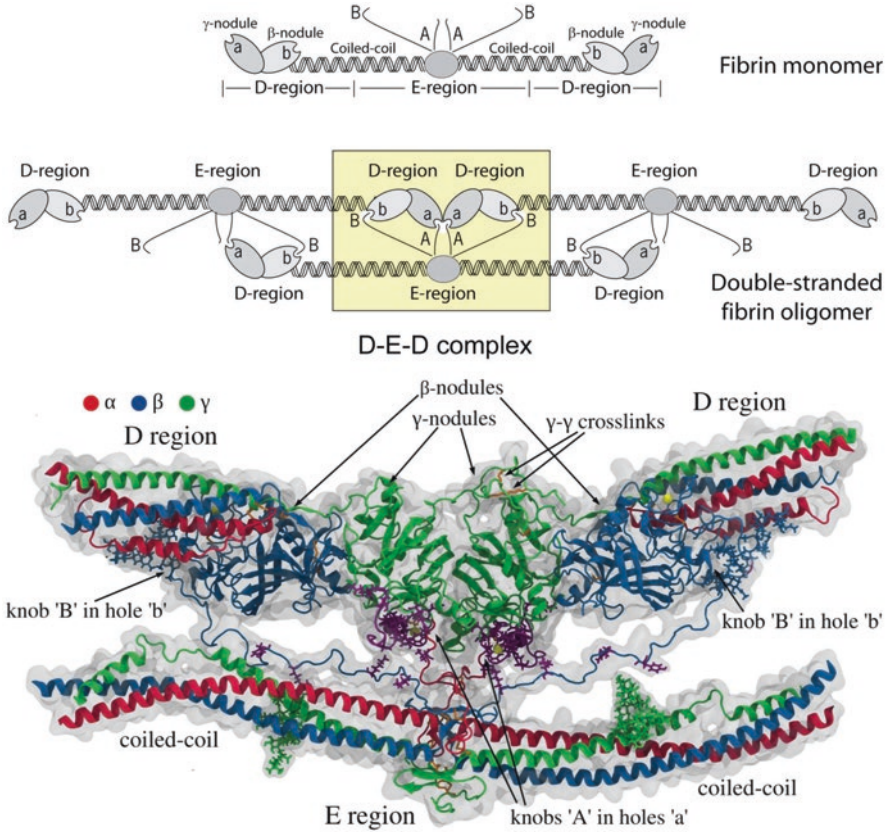
Fig. 13.4 Complementary binding sites or knob-hole interactions in fibrin polymerization. *Top*. Schematic diagram of knob-hole interactions. Knobs 'A' and 'B' in the central domain of a fibrin monomer are complementary to holes 'a' and 'b' that are always exposed at the ends of the protein. When the fibrinopeptides are removed by thrombin, exposing the knobs, knob-hole interactions occur, giving rise to the trimer shown and eventually to the two-stranded protofibril made up of half-staggered molecules. *Bottom*. Atomic resolution structure of the knob-hole interactions. The γ- and β-nodules near the ends of the molecule contain the holes 'a' and 'b', respectively, that are complementary to the knobs 'A' and 'B' in the central nodule. Most of these structures were derived from X-ray crystallographic data, although the disordered and/or flexible N-terminal regions of the α and β chains were derived from computational modeling (with permission from Elsevier Ltd.)

(Everse et al. 1998b; Kostelansky et al. 2002), evidence that these are holes 'a'. The hot spots of holes 'a' directly involved in the binding of the GPRP peptide were identified as residues γTrp315-Trp330, γTrp335-Asn365, and γPhe295-Thr305. However, using molecular dynamics simulations, it has been shown that during fibrin oligomerization the D:E interface includes binding sites beyond the 'A-a' knob-hole associations, confirming the existence of intermolecular binding sites 'A' and 'a' that are not limited to knobs 'A' and holes 'a' (Zhmurov et al. 2016).

Nevertheless, release of FpA and formation of knobs 'A' are necessary and sufficient to induce fibrin polymerization, resulting in formation of the so-called desA-fibrin. The existence of constitutively open holes 'a' in fibrinogen and fibrin is also essential for fibrin polymerization. If the holes 'a' are obstructed by the GPRP peptide (Everse et al. 1998b) or compromised by a replacement of the most important residue γAsp364 (Okumura et al. 1997), fibrin polymerization is abrogated. Overall, these and other data suggest that fibrin polymerization and clot formation are driven by the 'A-a' interactions.

13.3.4 Fibrin Oligomers and Protofibrils

Fibrin polymerization begins when two fibrin monomer molecules formed after cleavage of FpA interact to form a half-staggered dimer in which knob 'A' binds hole 'a', and there are two 'A-a' knob-hole interactions holding the two monomers together (Erickson and Fowler 1983). A third fibrin molecule added to a half-staggered dimer forms an end-to-end connection where the two adjacent molecules touch each other and the lateral D regions of two molecules form the D:D interface that comprises the junction between monomers in each of two strands in fibrin oligomers (Everse et al. 1998a; Fig. 13.4). The D:D interface comprises residues γ275–309 (Everse et al. 1998b) and is very weak, because it yields first upon forced stretching of fibrin(ogen) oligomers (Zhmurov et al. 2011). It was found that the D:D interactions involve residues γ275, γ308, and γ309 that are essential for elongation of fibrin strands (Marchi et al. 2006; Bowley et al. 2009). Fibrin monomers can add longitudinally to form longer two-stranded fibrin oligomers of varying length. Lateral interactions between two strands of fibrin oligomers are mediated by the central E region of one fibrin molecule and two lateral D regions of two other molecules (Fig. 13.4). The D-E-D complex is held together mainly by the 'A-a' knob-hole bonds and by additional interactions at the D:E and D:D interfaces (Kononova et al. 2013; Zhmurov et al. in preparation).

Fibrin monomers continue to add longitudinally to the oligomers, which lengthen further to make two-stranded protofibrils (Erickson and Fowler 1983), a critically important intermediate product of fibrin polymerization (Fig. 13.5). Protofibrils are usually about 0.5–0.6 μm in length, which corresponds to ~20–25 monomers, and they are long enough to self-interact and aggregate laterally (Erickson and Fowler 1983; Chernysh et al. 2011). Oligomers and protofibrils have been visualized by transmission electron microscopy (Erickson and Fowler 1983; Medved et al. 1990; Weisel et al. 1993; Chernysh et al. 2011; Huang et al. 2014), light microscopy (Chernysh and Weisel 2008; Chernysh et al. 2011), and atomic force microscopy (Yermolenko et al. 2011; Protopopova et al. 2015). In the presence of the fibrinogen γ′, which is a γ chain splice variant, protofibril formation is partially impaired by likely because of electrostatic repulsion (Cooper et al. 2003; Gersh et al. 2009; Allan et al. 2012).
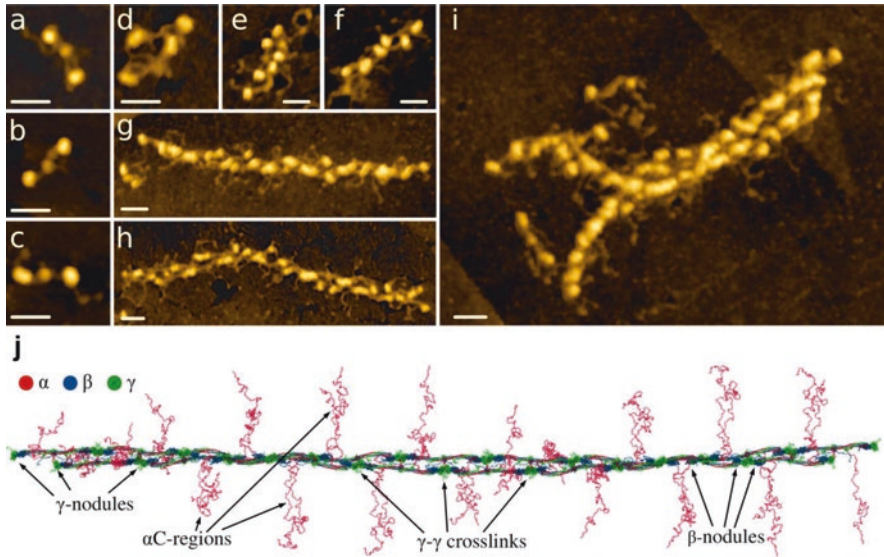
Fig. 13.5 Atomic force microscopy images of fibrinogen, fibrin oligomers, and protofibrils and reconstruction of a protofibril model. (**a–i**). Images by high-resolution atomic force microscopy (Published with permission and thanks to Drs. Anna D. Protopopova, Nikolay Barinov, Dmitry Klinov). All *magnification bars* = 30 nm. (**a–c**). Fibrin monomer with visible αC regions. (**d**). Fibrin dimer. (**e**). Fibrin trimer. (**f**). Fibrin tetramer. (**g–h**). Fibrin protofibrils. (**i**). Two protofibrils aggregating laterally. On the *left*, the two protofibrils are diverging, creating a branch-point. (**j**). Reconstruction of a twisted fibrin protofibril based on the X-ray crystallographic structure of fibrinogen (PDB Entry: 3GHG). The molecules are shown with addition of missing parts of the crystal structure reconstructed from molecular dynamics simulations, including the full-length αC regions. 'A-a' knob-hole bonds that are the major basis of fibrin polymerization are as in Fig. 13.4 (Published with permission and thanks to Dr. Artem Zhmurov)

13.3.5   Lateral Aggregation of Protofibrils

Protofibrils aggregate laterally to form more or less thick fibers, in which the half-staggered molecular packing gives rise to a 22.5-nm periodicity corresponding to half the length of the fibrin molecule (Fig. 13.3). This longitudinally ordered molecular packing of fibrin fibers was visualized by transmission electron microscopy of negatively contrasted specimens and atomic force microscopy as regular cross-striations (Weisel 1986; Yermolenko et al. 2011) (Fig. 13.10b). The mechanisms, structural motifs involved, and driving forces of the protofibrils' lateral aggregation are mostly still unknown. Importantly, protofibrils self-associate laterally only after they reach a certain threshold length, implying that the bonds mediating the interactions between protofibrils are weak and cooperative along the axis of the protofibril. At present, the structures that have been shown or presumed to participate in inter-protofibril lateral aggregation are the following: knobs 'B' and holes 'b', the αC regions, the C-terminal parts of the γ chains, adjacent β-nodules (Yang et al. 2000), the coiled-coils (Okumura et al. 2006), and N-glycans at residues βAsn364 and γAsn52 (Langer et al. 1988).

Because some other structural proteins that self-assemble have similar intermolecular interactions *in vitro* as *in vivo* (Weisel et al. 1978), useful mechanistic information about lateral aggregation of protofibrils has been gleaned from the interfaces formed between D regions in crystals. Thus, a mechanism for lateral aggregation has been proposed based on crystal packing via direct association of β-nodules of neighboring protofibrils, employing residues β330–375 (Yang et al. 2000). Some data suggest that lateral aggregation of protofibrils could be mediated by the residues βAla68 located in the N-terminal portion of the β chain (Mullin et al. 2000) and βGly15, the N-terminal residue of knob 'B', no matter whether FpB is released or not (Hirota-Kawadobora et al. 2003).

The packing in fibrin generally seems to be paracrystalline (Weisel et al. 1983; Weisel 1986), meaning that fibrin is more regularly organized axially than laterally, but there is also evidence for regularity in lateral packing (Torbet et al. 1981; Caracciolo et al. 2003; Yeromonahos et al. 2010). In addition, it appears that bundling of twisted protofibrils proceeds in such a way that the resulting fibers are also twisted (Medved et al. 1990). To maintain the 22.5-nm repeat, protofibrils that are newly added to the outside of a fiber must be stretched as their path length increases. This may comprise a thermodynamic mechanism to control the diameter of fibers, as the lateral aggregation stops when the protofibril stretching energy surpasses the energy of bonding.

An alternative model of fibrin formation and unusual structure is based on the formation of ultrathin fibrin sheets spanning channels on a plastic substrate, but the physiological relevance of these results is not yet known (O'Brien et al. 2008). Another proposed mechanism of the early stages of fibrin assembly implies that short thin fibrin branches form at an initial phase of polymerization, in which single-bonded "Y-ladder" polymers rapidly elongate before undergoing a delayed transition to the double-stranded fibrils (Rocco et al. 2014).

13.3.6 Role of 'B-b' Knob-Hole Interactions

When FpB is cleaved off, the β chain acquires a new N-terminal sequence Gly-His-Arg-Pro (GHRP), comprising knob 'B' capable of binding to hole 'b' located in the β-nodule (Fig. 13.4). The peptide GHRP, which reproduces the structure of knob 'B', has a much lower equilibrium binding affinity for fibrinogen (K_d = 140 µM) compared to the peptide GPRP (K_d = 25 µM) that mimics the structure of knob 'A' (Everse et al. 1998b). Moreover, a much larger piece of fibrin molecule from the central part (bearing knobs 'A') interacted with fibrinogen (bearing holes 'a') with even higher affinity (K_d = 5.8 ± 1.1 µM; Geer et al. 2007), suggesting that the 'A-a' interactions are not limited to the Gly-Pro-Arg motif. There must be additional interface beyond the 'A-a' bonds that has substantially higher binding strength, which has been confirmed using structural modeling (Kononova et al. 2013; Zhmurov et al. 2016).

Despite direct evidence for the existence of 'B-b' interactions (Litvinov et al. 2007a), the role of the 'B-b' interactions in fibrin formation is not quite clear because it is supported only indirectly. If fibrin is formed by cleavage of FpA only (while FpB remains uncleaved) the network consists of thinner fibers compared to the fibrin networks formed after cleavage of FpAs and FpBs., These data suggest that in the presence of 'B-b' interactions the fibers are thicker due to enhanced lateral aggregation of protofibrils (Blombäck et al. 1978), despite the fact that cleavage of FpB is not necessary for lateral aggregation and fiber formation. There are rare congenital homozygous fibrinogen mutations (fibrinogens Metz and Frankfurt XIII) in which FpB can be cleaved, while FpA cannot, that still form fibrin clots under the action of thrombin at low temperature (Galanakis et al. 1993). Thus, these clots are formed exclusively via 'B-b' interactions, indicating that they can actually occur in fibrin. Moreover, recombinant fibrinogen variants without functional holes 'a' (due to mutation of the γAsp364 residue) formed fibrin under the action of thrombin (that cleaved both FpAs and FpBs) but did not form fibrin under the action of reptilase (that cleaved FpAs only; Okumura et al. 2007). On the other hand, fibrinogen without functional hole 'b' (due to mutation BβAsp432Ala) formed normal fibrin (Bowley et al. 2009). Therefore, 'B-b' bonds not only exist physically but they form during fibrin polymerization under some special conditions (e.g., when 'A-a' bonds cannot form), but their role in normal fibrin formation remains largely undefined. Using molecular dynamics simulations, structural and thermodynamic characteristics of the 'B-b' interactions has been gleaned (Kononova et al. 2013). It has been proposed that 'B-b' interactions have effects on the susceptibility of clots to proteolytic digestion (Doolittle and Pandi 2006). Besides the known 'A-a' and 'B-b' knob-hole interactions, experimental data suggest that the 'A:b' interactions are physically possible, while the 'B:a' interactions are unlikely to exist (Litvinov et al. 2005, 2007a). Fibrin structure and properties are affected by addition of a knob 'B' mimetic combined with polyethyleneglycol, thus indirectly confirming the functional importance of 'B-b' knob-hole interactions (Brown et al. 2015).

13.3.7 The Role of the αC Regions in Fibrin Formation

During fibrin polymerization, the protruding and flexible αC regions (see Sects. 13.2.4 and 13.2.5) can self-interact both within and between protofibrils; these αC-αC interactions can lead to formation of αC polymers that are reinforced with additional crosslinking by Factor XIIIa (Tsurupa et al. 2011; Fig. 13.3). The αC-αC polymerization involves two mechanisms. One is self-association of the αC-domains that occurs via their N-terminal subdomains by β-hairpin swapping. The second is the interaction of the C-terminal subdomain with the αC-connector (Tsurupa et al. 2012). Although they are not necessary, the αC regions are known to augment lateral aggregation (Weisel and Medved 2001; Litvinov et al. 2007b; Tsurupa et al. 2011). Fibrinogen lacking the αC regions forms clots with thinner fibers and a higher density of branch points than clots made of full-length fibrinogen (Collet

et al. 2005). Furthermore, a recombinant fibrinogen construct that has the longer human αC regions replaced with the shorter chicken αC regions displayed impaired lateral aggregation of protofibrils (Ping et al. 2011).

13.3.8 Fibrin Branching and Network Architecture

As fibrin fibers thicken by lateral aggregation and grow in length, they also branch, yielding a space-filling 3D network. Electron microscopy revealed at least two different molecular mechanisms by which branch points may form (Figs. 13.3 and 13.5i). One of them known as a "bilateral junction", originates from two protofibrils that undergo incomplete lateral aggregation but diverge into two separate protofibrils, each of them giving rise to a new fiber (Mosesson et al. 1993). The second type of branchpoint, called a "trimolecular junction" or "equilateral junction", forms when a fibrin monomer attached to the end of a protofibril via only one 'A-a' bond (or one γ-nodule), such that both the monomeric molecule and the protofibril to which it is bound can grow independently, forming two strands each (Fogelson and Keener 2010). In either case, most of branch points in clots consist of three fibers of about the same diameters joined together (Ryan et al. 1999), suggesting that the type of initial branchpoint does not affect much the final network structure. Branching has been observed directly in an imaging study of the early stages of fibrin polymerization (Chernysh et al. 2011). Finally, an increase in the number of branch points in a clot normally correlated with a decrease of the fiber diameters (Ryan et al. 1999). The aggregate of data support a notion that that branching (as a part of elongation process) and lateral aggregation compete. In other words, the prevalence of lateral aggregation leads to fibrin with thicker fibers and fewer branch points, while reduced lateral aggregation will result in formation of fibrin with thinner fibers and more branch points.

13.3.9 Fibrin Structure and the Gelation Point

Formation of a fibrin clot is a transition from sol to gel upon formation of a three-dimensional filamentous network (Fig. 13.6). The gelation point determined as blood/plasma clotting time is commonly used in clinical assays as a test to reveal coagulation disorders. The gelation point occurs when only about 15–20 % of the fibrinogen is converted to fibrin, which is enough to form the gel (Chernysh and Weisel 2008). Some correlations between gelation time and final clot structure have been established (Blombäck and Okada 1982), but the formation of a stable fully branched network is not yet completed at the gelation point, so that new fibers and branch points continue to form in the gel (Chernysh and Weisel 2008). The structure of fibrin networks can be determined and quantified using scanning electron microscopy by a number of parameters, such as the fiber diameter, fiber density, number of

Fig. 13.6 Fibrin clot network. A 3-dimensional reconstruction of a hydrated fibrin gel obtained using fluorescent confocal microscopy. Fibers are very straight under tension and branch to form a network. Fibrin(ogen) was fluorescently labeled with Alexa 488 (Brown et al. 2009)
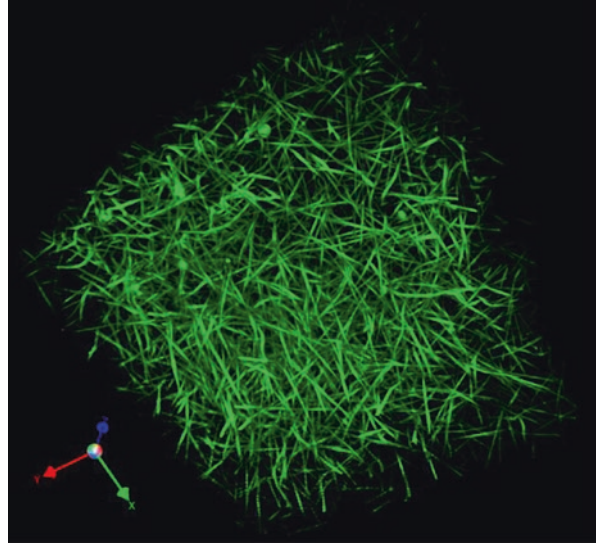
branch points, fiber length, and the size of the pores (Fig. 13.10a). All of these parameters are variable and depend on the kinetics of fibrin polymerization. Various biophysical techniques have been applied to study the fine structure of fibrin clots that revealed the complex structural hierarchy at different spatial scales (Ryan et al. 1999; Ferri et al. 2002; Guthold et al. 2004; Evans et al. 2010; Yeromonahos et al. 2010; Magatti et al. 2013). Remarkably, the diffusion rate of proteins in fibrin clots does not depend much on the structure of the network because there are large pores. However, permeability of fibrin clots for fluid (Okada and Blombäck 1983) or nanoparticles (Spero et al. 2011) perfused through the clot depends strongly on the pore size and hence on overall clot structure. The diffusivity of the fibrin network is reduced dramatically when it is embedded with blood cells and compressed, as happens in a contracted whole blood clot (Cines et al. 2014).

13.3.10   Factor XIIIa-Catalyzed Covalent Crosslinking of Fibrin

To stabilize the clot against proteolytic and mechanical insults, fibrin is covalently crosslinked by the plasma transglutaminase, Factor XIIIa, an active from of Factor XIII zymogen activated by thrombin in the presence of Ca^{2+}. The C-terminal ends of the γ chains of fibrin(ogen) have amino acid residues comprising a crosslinking site for two end-to-end interacting molecules that form covalent isopeptide ε-(γ-glutamyl)-lysyl bonds between the γLys406 of one molecule and γGln398/399 of another molecule (Fig. 13.7). There has been a disagreement about longitudinal γ-γ-crosslinking within a strand of a protofibril versus transverse γ-γ-crosslinking

Fig. 13.7 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

$$\boxed{Fibrin}-(CH_2)_2-C\overset{O}{\underset{NH_2}{\big\langle}} + NH_2-(CH_2)_4-\boxed{Fibrin}$$

Glutamynyl   Lysyl

FXIIIa (transglutaminase)

NH_3

$$\boxed{Fibrin}-(CH_2)_2-\overset{O}{\overset{\|}{C}}-\underset{H}{\overset{}{N}}-(CH_2)_4-\boxed{Fibrin}$$

ε-(γ-glutamyl)-lysyl bond
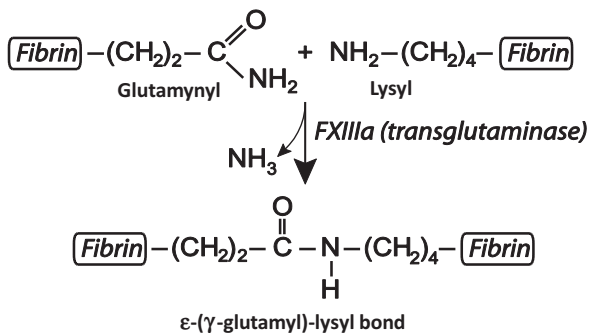
between strands (Weisel 2004; Mosesson 2004), but new evidence has been provided for the longitudinal orientation of these bonds (Rosenfeld et al. 2015). Formation of the same isopeptide bonds is catalyzed at a smaller rate between the αC regions to stabilize long αC polymers (Matsuka et al. 1996). Crosslinking also occurs among α and γ chains followed by formation of α-γ-heterodimers (Standeven et al. 2007). Before crosslinking, fibrin polymerization is reversible, with fibrin being an equilibrium polymer (Chernysh et al. 2012). After crosslinking within and between protofibrils, polymerization becomes irreversible, and the clot is more stable, mechanically strong, and resistant to fibrinolysis. The crosslinked fibrin can be dissolved either by reduction of disulphide bonds that hold polypeptide chains together or by chemical/enzymatic hydrolysis of peptide bonds. A normal genetic variant of Factor XIII with a Val34Leu polymorphism forms either porous permeable clots or dense clots with reduced permeability at various fibrinogen levels (Lim et al. 2003). A synthetic hemostatic polymer has been recently designed that stabilizes fibrin clots chemically, thus mimicking the crosslinking effect of Factor XIIIa (Chan et al. 2015).

13.4 Variations and Modulation of Fibrin(ogen) Structure and Properties

13.4.1 Genetic Polymorphisms of Fibrinogen

Variants of fibrinogen are present in the blood as a result of several common polymorphisms or normal alternative primary structures. The most abundant fibrinogen variants contain two types of γ chains, γA and γ′, that result from an alternative polyadenylation signal in intron 9 of the *FGG* gene (Francis et al. 1980). About 8–15 % of total fibrinogen contains the γ′ chain, of which the majority is in the heterodimeric γA/γ′ form with a homodimer γ′/γ′ comprising only about 1 % (Chung and Davie 1984). The presence of γ′ chains was shown to slow down lateral aggregation of protofibrils and alter fibrin formation and structure (Ajjan et al. 2009;

Domingues et al. 2015; Muthard et al. 2015). An increase in plasma fibrinogen γ′ concentration is associated with the risk of myocardial infarction and other thrombotic states (Mannila et al. 2007; Lovely et al. 2002; Uitte de Willige et al. 2005, 2009). Fibrinogen polymorphisms with functional consequences can occur in the polypeptide chains other than the γ chain. αE fibrinogen (see Sect. 13.2.1) has a reduced rate of fibrin polymerization and forms thinner and more branched fibers than fibrin networks containing the α chains (Mosesson et al. 2004). The fibrinogen Aα chain gene *FGA* polymorphism 2224G/A has been associated with reduced clot permeability (Mannila et al. 2006), while the Ala312 allele of *FGA* 6534/Thr312Ala was associated with increased clot stiffness (Standeven et al. 2003), and the Lys448 allele of the fibrinogen Bβ chain gene (*FGB*) polymorphism BβArg448Lys resulted in a compact fibrin network structure resistant to lysis (Ajjan et al. 2008). The other most widespread polymorphisms in the fibrinogen genes occur in the noncoding regions and can result in changes in plasma fibrinogen levels. There are many more examples of strong associations between fibrinogen polymorphisms, clot structure and properties, and disease (Ariëns et al. 2002; Scott et al. 2004).

13.4.2 *Post-translational Modifications and Heterogeneity of Fibrinogen*

There are many molecular forms of fibrinogen present in blood, as originally detected from variations in biochemical properties and gel electrophoretic behavior. It has been estimated that fibrinogen may occur in more than a million non-identical forms in a healthy individual as a result of the many combinations of modified or inherently polymorphic sites (Henschen-Edman 2001). Several of these genetically determined variations have already been mentioned, but there are also post-translational heterogeneities originating from multiple biochemical reactions that accompany various physiological and especially pathological conditions, such as inflammation or ischemia. These reactions can modify the fibrinogen molecule in many ways, such as phosphorylation at specific seryl and threonyl sites, prolyl hydroxylation, tyrosyl sulfation, asparaginyl or glutaminyl deamidation, N-terminal pyroglutamate formation from glutaminyl precursors, oxidation of methionine, histidine and tryptophan residues, tyrosine nitration, modifications of cysteine residues, formation of dityrosine and carbonyl groups, etc. The C-terminal portion of the Aα chain is especially susceptible to limited cleavage by intracellular and extracellular proteolytic enzymes, but some digestion also occurs at specific sites in the Bβ and γ chains, such that lower molecular weight forms are commonly present in plasma fibrinogen. Alternative N-glycosylation can be another source of fibrinogen heterogeneity because it may result in formation of oligosaccharides with a variable structure (Brennan 2015). Acquired abnormal fibrinogen variants occur in patients with several conditions via non-enzymatic reactions, including glycation of lysine residues in uncompensated diabetes mellitus (Dunn and Ariens 2004) or

homocysteinylation in hyperhomocysteinemia (Sauls et al. 2003). Fibrinogen derivatives can form in pathological conditions associated with thrombin generation, activation of fibrinolytic enzymes, or immune reactions, for example crosslinked fibrin(ogen) degradation products form when fibrinolytic activity in the blood is excessive, and antibody-fibrinogen complexes are formed in some autoimmune diseases. Individuals treated with acetylsalicylic acid (aspirin) have acetylated lysine residues (Ajjan et al. 2009). Oxidative stress has been widely implicated in physiological processes such as aging, in various disease pathogenesis, including arterial and venous thrombosis. Proteins are major targets for oxidants, and fibrinogen is a common target for oxidative post-translational modifications (Martinez et al. 2013; Rosenfeld et al. 2014). Neutrophils and monocytes, the most prevalent leukocytes in the sites of inflammation and venous thrombi, generate nitrating metabolic intermediates capable of nitration of fibrinogen (Heffron et al. 2009; Martinez et al. 2012). Tobacco smokers tend to have nitration of tyrosine and oxidation of methionine, histidine or tryptophan residues (Parastatidis et al. 2008). Many of these chemically modified forms are associated with differences in functional and structural properties of fibrinogen and fibrin, including a thrombogenic phenotype associated with increased risk of arterial and venous thrombosis (Nowak et al. 2007; Parastatidis et al. 2008; Paton et al. 2010; Sauls et al. 2006; Undas et al. 2006; Weigandt et al. 2012).

13.4.3 Hereditary Fibrinogen Defects (Dysfibrinogenemias, Afibrinogenemia, and Hypofibrinogenemia)

Dysfibrinogenemias are characterized by inherent structural changes in the fibrinogen molecule that commonly result in alterations in clotting or other functional aspects of the protein (Casini et al. 2016). Traditionally, particular defective fibrinogen variants are named after the city of their discovery or where the patient lived. Most congenital defects are rare but have led to important insights into fibrin(ogen) structure-function relationships. Many of these mutations have been described in reviews (Galanakis 1993; Roberts et al. 2001; Matsuda and Sugo 2001; de Moerloose and Neerman-Arbez 2009; Asselta et al. 2006; de Moerloose et al. 2013; Asselta et al. 2015; Casini et al. 2015). In addition, the website, http://www.geht.org/databaseang/fibrinogen, maintains a database of fibrinogen mutations and their functional consequences.

Mutations that give rise to dysfibrinogenemias are commonly caused by single base mutations that lead to the substitution of a single amino acid residue, but other mutations can give rise to a stop codon, resulting in a truncation of one of the chains. In addition, base additions or deletions may occur, with consequences for fibrinogen structure and function. Such mutations can cause a predisposition to thrombosis, or bleeding, or be asymptomatic, depending on their functional effects. They can affect fibrinopeptide cleavage, fibrin polymerization, Factor XIIIa-catalyzed crosslinking,

susceptibility to fibrinolysis or integrin αIIbβ3-mediated platelet aggregation. In homozygous forms of the dysfibrinogenemias, the mutations occur in all fibrinogen molecules so only abnormal homodimers are present in the blood, but these mutations are rare, so most dysfibrinogenemias are heterozygous. As a result, the pool of fibrinogen molecules in most subjects with a dysfibrinogenemia consists of various proportions of normal homodimers, mutant homodimers and heterodimers.

Most dysfibrinogenemias are likely to be undetected by clotting assays, but some of the most commonly found alterations are in the N-terminal part of the Aα chain, in the regions responsible for thrombin binding or cleavage or in or near knob 'A', i.e. the AαGly17-Pro18-Arg19 motif. One such mutation is a substitution of AαArg16 (the last residue of FpA) by His or Cys, the former causing delayed release of FpA, with the latter completely abrogating its release. With no FpA cleavage, FpB is released slowly and clots made of desB-fibrin form, but only at lower temperatures (Shainoff and Dardik 1979). Mutations of AαGly17, AαPro18, AαArg19 or AαVal20 result in defective polymerization because of alterations of knobs 'A.'

Mutations of the Bβ chain are less common. Substitution of Cys for BβGly at position 15 results in slow polymerization from delayed release of FpB (Sugo et al. 2000). Mutation of BβAla68 to Thr results in defective binding of thrombin to fibrin and consequent thrombosis (Koopman et al. 1992), since fibrin is a physiological absorbent of thrombin and was initially called antithrombin I (Mosesson 2007).

In the γ chain, many mutations in the C-terminal portion have been identified, particularly at position γArg275 (Cote et al. 1998). This residue makes contacts at the D:D interface, so that mutations affect fibrin polymerization. Mutations in or near the hole 'a' are common but patients with most of these substitutions are asymptomatic, presumably because they are heterozygous.

Some point mutations are responsible for congenital hypo- and afibrinogenemia, as a result of defects in molecular processing, assembly, secretion, and domain stability of fibrinogen. Afibrinogenemia is an autosomal recessive disorder characterized by the complete absence of detectable fibrinogen in the blood and hypofibrinogenemia is characterized by a reduced level of fibrinogen. Analysis of the three fibrinogen genes in affected individuals has led to the identification of several causative mutations (Brennan et al. 2001; Neerman-Arbez 2001), mostly as a result of mutations in the *FBA* gene. The analysis of the degree of severity of the fibrinogen disorders associated with truncation of the Aα chain suggest that formation of the distal disulfide ring of the coiled-coil is essential for assembly and secretion of fibrinogen molecules.

13.4.4   *Environmental Conditions of Fibrin Formation*

Fibrin formation, structure, and properties both *in vitro* and *in vivo* are strongly affected by external factors, such as ionic strength, composition, pH and various endogenous and exogenous substances, e.g., polyphosphate, mono-, oligo- and polysaccharides, peptides, lipids, proteins, nucleic acids, medications, as well as many other normal and pathological, natural and artificial compounds present in

blood and injured tissues. Thrombin activity in blood has a profound effect on fibrin, with a high thrombin activity resulting in clots with thinner fibers, a higher density of branch points, and smaller pores, while a low thrombin activity results in thicker fibers with fewer branch points and larger pores. Most of these structural variations are based on the kinetics of individual steps of fibrin polymerization (Weisel and Nagaswami 1992). The structure of fibrin networks is often affected by physical factors, such as hydrodynamic flow or a strong magnetic field that result in formation of oriented anisotropic fibrin fibers (Gersh et al. 2010; Campbell et al. 2010). Fibrin structure and properties are greatly influenced by the presence of blood cells, namely activated platelets and erythrocytes (Aleman et al. 2014; Malecki et al. 2015) that form a natural and very active environment for clot formation. In addition to whole cells, the effects of circulating cell-derived microparticles on the fibrin clot structure and properties have been recently demonstrated (Zubairova et al. 2015). Fibrin can interact with other components of the extracellular matrix, both filamentous and non-filamentous, that not only affects fibrin structure but imparts additional mechanical and chemical stability (Maquart and Monboisse 2014). The structure of fibrin clots can be directly related to clinical conditions associated with thrombosis (Collet et al. 2006; Zalewski et al. 2015).

Here is a list of plasma proteins that bind specifically to fibrinogen or fibrin or both, with some indications of the biological significance:

Actin	modulation of clot structure and properties (Talens et al. 2012)
Albumin	modulation of lateral aggregation and clot structure (Galanakis et al. 1987; Talens et al. 2012)
α1-Antitrypsin	local regulation of proteases involved in coagulation or fibrinolysis (Talens et al. 2012)
α2-Antiplasmin	local regulation of fibrinolysis (Tsurupa et al. 2010)
Apolipoproteins A-IV, A-I, J, E	associated with high-density lipoproteins; the physio logic role of binding to fibrin(ogen) is not clear (Talens et al. 2012)
Carboxypeptidase N	a local fibrinolysis inhibitor (Talens et al. 2012)
CD44	mediates tumor metastasis at the sites of fibrin deposition (Alves et al. 2009)
Coagulation Factor Xa	a negative feedback for thrombin generation (Iino et al. 1995)
Coagulation Factor VIII	platelet-attached soluble fibrin mediates binding of Factor VIII (Gilbert et al. 2015)
Coagulation Factor XIII	crosslinking of fibrin and multiple plasma proteins (Weisel and Dempfle 2013)
Complement C3	delayed fibrinolysis (Howes et al. 2012)

Factor H-related proteins-associated lipoprotein particles (FALP)	the physiologic role of fibrinogen on FALP is not clear (Park and Wright 2000)
Ferritin	some fibrinogens circulate in the form of a complex with ferritin (Takahashi et al. 2013)
Fibroblast Growth Factor-2	augmented angiogenesis and cell proliferation at the sites of fibrin deposition (Sahni et al. 1999)
Fibulin	interferes with the fibrin assembly (Tran et al. 1995)
Haptoglobin	associated with high-density lipoproteins; the physiologic role of binding to fibrin(ogen) is not clear (Talens et al. 2012)
Hepatocyte-derived fibrinogen related protein-1	liver cell growth regulation (Talens et al. 2012)
Immunoglobulins	modulate fibrin structure, localize immune and inflammatory reactions (Talens et al. 2012)
Interleukin-1β	modulation of inflammation (Sahni et al. 2004)
Lipoprotein(a)	modulation of fibrinolysis (Weisel and Litvinov 2014)
α2-Macroglobulin	local regulation of proteases involved in coagulation or fibrinolysis (Talens et al. 2012)
Myosin	modulation of fibrinolysis (Kolev et al. 2003)
Plasma fibronectin	platelet adhesion and thrombus formation (Gailit and Ruoslahti 1988; Talens et al. 2012)
Plasminogen	promotion of fibrinolysis (Weisel and Litvinov 2014)
Plasminogen activator inhibitor-1	modulation of fibrinolysis (Smolarczyk et al. 2005)
Serum amyloid P	associated with high-density lipoproteins; the physiologic role of binding to fibrin(ogen) is not clear (Talens et al. 2012)
Thrombin	elimination of active thrombin from the circulation (Weisel and Dempfle 2013)
Thrombin-activatable fibrinolysis inhibitor (TAFI)	modulation of fibrinolysis (Valnickova and Enghild 1998)
Thrombospondin	modulation of fibrinolysis (Bacon-Baguley et al. 1990)
Tissue-type plasminogen activator (t-PA)	promotion of fibrinolysis (Medved et al. 2001)
Vascular Endothelial Growth Factor (VEGF)	promotion of angiogenesis at the sites of fibrin deposition (Sahni and Francis 2000)
VE-cadherin	suppression of inflammation by inhibiting leukocyte transmigration at the sites of fibrin deposition (Yakovlev et al. 2011)

Very low density lipoprotein receptor (VLDLR)	promotion of inflammation and transendothelial migration of leukocytes at the sites of fibrin deposition (Yakovlev and Medved 2015)
Vitronectin	platelet adhesion and thrombus formation (Schvartz et al. 2002; Podor et al. 2002)
von Willebrand Factor	platelet adhesion and thrombus formation (Miszta et al. 2014)

13.4.5 Fibrin Formation Under Hydrodynamic Flow

Blood flow is one of the most important physical factors that affects profoundly formation of the fibrin network, its structure and properties *in vivo* (Swieringa et al. 2016). Clot formation in static conditions stops when all soluble fibrinogen is converted to insoluble fibrin. Therefore, in normal human plasma with a fibrinogen concentration of about 3 g/L, a fibrin clot will consist of only 0.3 % protein and 99.7 % liquid by mass. Alternatively, under flow conditions more fibrinogen is added to the forming clot, so the clot will contain more protein and have a different structure with much denser, thicker and bundled fibers (Silvain et al. 2011; Neeves et al. 2010). Fibrin formed under flow conditions also has some fibrin fibers orientated along the direction of flow (Gersh et al. 2010; Campbell et al. 2010; Whittaker and Przyklenk 2009), which changes clot mechanical properties and affects its susceptibility to enzymatic lysis (Campbell et al. 2010; Varju et al. 2011). Furthermore, it has been proposed from microfluidics experiments that the shear forces of blood flow determine the likelihood of embolization, namely the rupture of a piece of a clot that is carried by the blood stream to block another vessel (Colace et al. 2012; Brass and Diamond 2016).

13.5 Fibrin Mechanical Properties and Their Structural Origins

13.5.1 General Remarks

Fibrin mechanics is an important and rapidly developing field because fibrin networks form at sites of vascular injuries and perform a mechanical task of stemming blood flow by forming a gel under hydrodynamic blood shear (Campbell et al. 2010; Flamm and Diamond 2012), contraction of platelets (Lam et al. 2011), contraction of adjacent muscles, and pulsation of a vessel wall (Gasser et al. 2008; Ashton et al. 2009). Therefore, the outcomes of many bleeding and thrombotic disorders, including thromboembolism, are largely determined by the mechanical behavior of fibrin networks (Tran et al. 2013). In addition, fibrin has been used as a biomaterial for

numerous purposes of surgical repairs and tissue engineering to stop or control bleeding or to form a provisional fibrin matrix for growing blood vessels and tissue regeneration, all of which strongly depend on the mechanical properties of fibrin gels.

13.5.2 Viscoelastic Properties of Fibrin

Fibrin is a viscoelastic polymer, which means that it has both elastic and viscous properties. The elasticity (or stiffness) is characterized by reversible mechanical deformation, while viscosity (or plasticity) is characterized by a slow irreversible deformation (creep) induced by force. Viscoelastic biomaterials differ in the relative degrees of both elastic and viscous properties. The elastic response of the fibrin clot is characterized by the shear storage modulus, G′, corresponding to the part of shear stress that is in phase with strain. The viscous response of the clot to applied shear is measured by the shear loss modulus, G″, where strain lags stress. The storage and loss moduli determine how the clot responds to the forces to which it is subjected. The ratio G″/G′ is often used to characterize relative viscosity and stiffness of a fibrin clot.

Clots derived from the blood of subjects with pulmonary embolism showed accelerated establishment of viscoelastic properties compared to healthy donors (Martinez et al. 2014). In addition, the stiffness of clots formed from the blood of patients who have had heart attacks at an early age is 50 % greater than that of controls (Collet et al. 2006). The clot fractal dimension, based on viscoelastic properties of incipient blood clots, has been used as a biomarker of prothrombotic clot microstructure (Lawrence et al. 2015; Davies et al. 2015). Platelets sense the stiffness of the underlying fibrin/fibrinogen substrate so that higher substrate stiffness leads to increased platelet activation, adhesion and spreading (Qiu et al. 2014; Wufsus et al. 2015). Lastly, the stiffness of the fibrin scaffold of occlusive thrombi is a major determinant for effectiveness of their mechanical damage and removal to restore the impaired blood flow (Weiss et al. 2013).

13.5.3 Non-linear Elasticity and High Extensibility of Fibrin

The elasticity of fibrin clots is generally characterized by a stress-strain curve, in which an applied stress (force/area) is plotted against the degree of induced deformation (strain) (Fig. 13.8). At low strains or deformations of fibrin, stress is directly proportional to strain and the slope of the curve (the elastic modulus) is constant. At larger strains, the linearity is broken and the slope of the curve increases dramatically, so that the elastic modulus or stiffness of the clot increases up to one order of magnitude or more (Janmey et al. 1983). This non-linearity is called strain hardening or strain stiffening and is a fundamental property of biological gel-like structures (Storm et al. 2005).
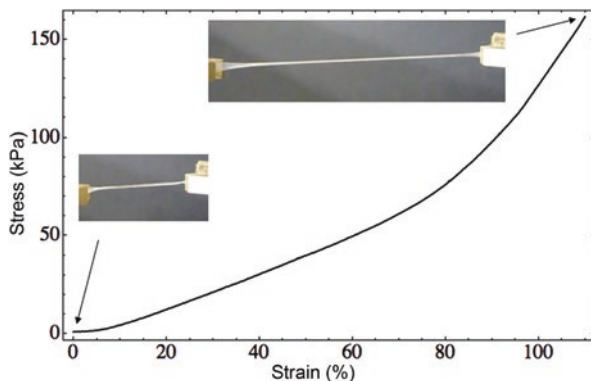
Fig. 13.8 Stress-strain curve of a fibrin clot. Representative stress-strain curve of a cylindrical fibrin clot reaching greater than a two-fold longitudinal stretch. As the strain increases, the stress on the clot increases linearly until a strain of ~80 % is reached, at which point the sample hardens and enters a new regime with a steeper slope, corresponding to increased stiffness or strain hardening or stiffening. *Insets show photographs* of the initial clot and stretched clot

Fibrin is a highly extensible polymer, which means that under stress blood clots will tend to stretch rather than break. Plasma clots stabilized with Factor XIIIa could be stretched to over three times their relaxed length before breaking (Brown et al. 2009). The extraordinary extensibility, unusual visoelasticity, including strain stiffening, has been demonstrated and quantified at the level of individual fibers. This can be observed when a fibrin fiber is laterally stretched with a tip of an atomic force microscope, so that crosslinked and uncrosslinked fibrin fibers are stretched to about 2.5 and 3.3 times their original length before rupturing (Liu et al. 2006, 2010; Guthold et al. 2007). The propagation of strain-stiffening throughout the entire fibrin gel is largely determined by the fine structure and non-linear elastic properties of individual fibrin fibers (Hudson et al. 2015; Piechocka et al. 2010). As fibers are stretched, they become stiffer than any surrounding fibers at lower strains, which allow the more strained, stiffer fibers to distribute the strain load to the less strained fibers and reduce strain concentrations (Hudson et al. 2010). In addition to shear and tension, the non-linear elasticity of fibrin has been observed also in response to compressive deformations (Kim et al. 2014).

There are a number of environmental factors that modulate fibrin stiffness both *in vitro* and *in vivo*. The main physiological modulator of fibrin mechanics is Factor XIIIa, which catalyzes fibrin crosslinking (Fig. 13.7) and increases the elastic modulus of fibrin several-fold, apparently by fiber compaction (Kurniawan et al. 2014). Based on the *in vitro* effects of zinc that reduces fibrin clot stiffness, it was suggested that zinc released from activated platelets may modulate clot strength and stability in cooperation with Factor XIIIa. Fibrin elasticity can be modulated by other physical influences (Kotlarchyk et al. 2011; Munster et al. 2013) or biochemical modifications as well as by blood components and cells incorporated into the

fibrin network (Rojas et al. 2009; Weigandt et al. 2012; Lauricella et al. 2013; Jansen et al. 2013; Henderson et al. 2015).

The tunable non-linear elasticity of fibrin may be important biologically because it allows fibrin clots to be compliant at smaller strains and then become stiffer at larger deformations that could otherwise threaten clot integrity and make them prone to embolization. In addition, the complex mechanical behavior of fibrin is important for the interaction between cells and extracellular matrix (Wen and Janmey 2013). Since mechanical stress makes fibrin more resistant to fibrinolysis (Varju et al. 2011; Bucay et al. 2015), fibrin elasticity may be a significant determinant of susceptibility of clots and thrombi to enzymatic lysis (Rottenberger et al. 2013; Longstaff et al. 2013).

13.5.4 *Multiscale Structural Mechanics of Fibrin Clots*

It has been shown that fibrin mechanics are governed by a structural hierarchy, implying that fibrin deformation is accompanied by multiple structural rearrangements at different scales, namely the molecular level, individual fibers, fiber network, and the whole clot (Brown et al. 2009; Purohit et al. 2011; Piechocka et al. 2010; Yeromonahos et al. 2010).

At the macroscopic scale (10^{-2} m), in addition to their large extensibility, fibrin clots also display a dramatic decrease in volume when they are stretched (Brown et al. 2009). The shrinkage of the stretched clot is due to water expulsion and network densification, as confirmed by an approximately tenfold increase in the protein content in clots stretched threefold. This observation might be related to the phenomenon of negative normal stress observed for networks of semiflexible polymers (Guo et al. 2009). An alternative explanation is that the volume change is associated with a molecular structural transition that occurs in stretched fibrin fibers (Brown et al. 2009; Purohit et al. 2011).

At the network scale (10^{-5} m), when strain is applied, the fibers begin to align along the direction of strain (Brown et al. 2009; Litvinov et al. 2012; Purohit et al. 2011), the fibers become thinner, closer together, and bundle (Müller et al. 1984; Brown et al. 2009). A dramatic rearrangement of the fibrin network was also observed in response to compressive deformation (Kim et al. 2014): the fibrin network density increased, fibers reoriented in the compression plane and shorter fiber segments were formed as a result of fiber crisscrossing. These transformations of the network structure are accompanied by a dramatic increase of fibrin elasticity.

At the fiber scale (10^{-6} m), in response to compression or shear, individual fibrin fibers in the network begin to buckle and bend in the direction of deformation (Lindstrom et al. 2013; Kim et al. 2014). Because buckling and subsequent bending make fibers more compliant, the stiffness of the network in the transverse shear direction gradually decreases with compression.

13.5.5 Molecular Structural Origins of Fibrin Mechanical Properties

At the molecular level, fibrin has the potential to elongate more than 5-fold, from 45 nm to about 250 nm (Fig. 13.9). The most important structural changes during fibrin deformations include unfolding of coiled-coils (Brown et al. 2007; Lim et al. 2008; Litvinov et al. 2012), the γ-nodules (Averett et al. 2008, 2009), and extension of the αC regions. In addition, other studies point to a role of network branch points (Carlisle et al. 2010) and the γA/γ´ splice variants (Gersh et al. 2009; Allan et al. 2012). Based on the known crystal structure of a folded fibrin(ogen) molecule (Fig. 13.9), it could be predicted that full hypothetical unfolding of compact structures would result in a ~4.7-fold elongation (Zhmurov et al. 2011; Hudson et al. 2013), which characterizes a great potential contribution of molecular unfolding in large deformations of fibrin.

The first direct experimental observation for an α-helix to β-strand conversion of the coiled-coils accompanying extension of fibrin was wide-angle X-ray scattering of squeezed fibrin films (Bailey et al. 1943). Unlike the extensively studied α-helix to β-strand transition in keratin, this preliminary result was not pursued until
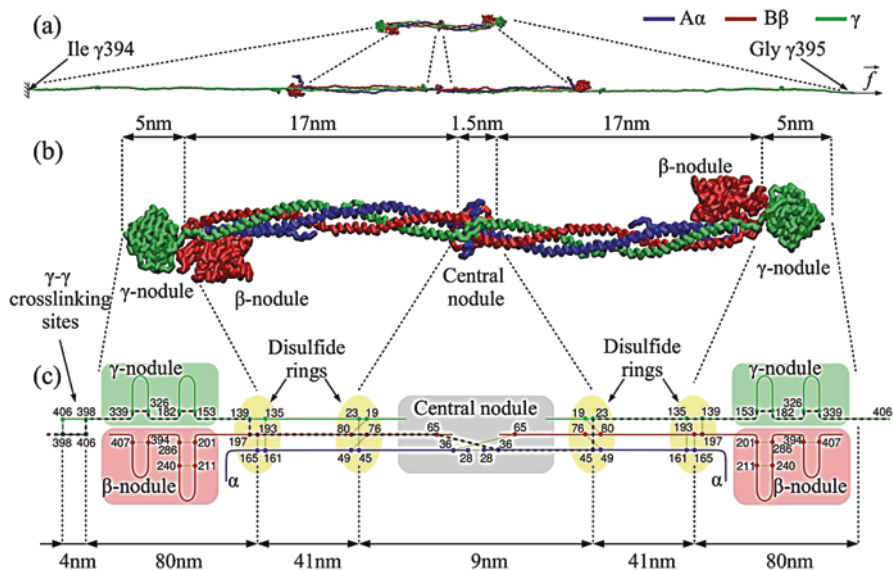
Fig. 13.9 Unfolding of fibrin(ogen). A schematic representation of the fibrin(ogen) molecule in the naturally folded (**b**) and fully unfolded states ((**a**) and (**c**)). The molecule is constrained at the C-terminal part of one γ chain, and mechanical force is applied to the C-terminal part of the other γ chain. (**a**) Shows full unfolding without much detail, while the structural details are given in (**c**), showing schematically the lengths of the central nodule, γ-nodules, β-nodules, coiled-coils, taking into account the disulfide bonds. Dimensions are shown in the compact crystal structure (**b**), and the contour lengths of various structural elements are shown in the fully unfolded state (**c**), assuming a contour length per residue of 0.38 nm (Zhmurov et al. 2011, with permission of Elsevier Ltd.)

Fig. 13.10 Unfolding of the coiled-coils of fibrin. (**a**). scanning electron micrograph of a fibrin clot, with a box enclosing part of a fiber. (**b**). A transmission electron micrograph of a negatively contrasted fibrin fiber showing the ultrastructure, with the 22.5 nm repeat arising from the half-staggering of 45-nm molecules. (**c**). A fibrin trimer from X-ray crystallographic data and molecular dynamics simulations of regions not present in the crystal structure. (**d**). Fibrin α-helical coiled coils undergoing a forced transition from α-helix to β-sheet. The mechanical transition from α-helical coiled coils to β-sheets in the fibrin(ogen) molecule was characterized using molecular dynamics simulations of their forced elongation and theoretical modeling (Adapted with permission from Zhmurov et al. 2012. Copyright 2012 American Chemical Society)
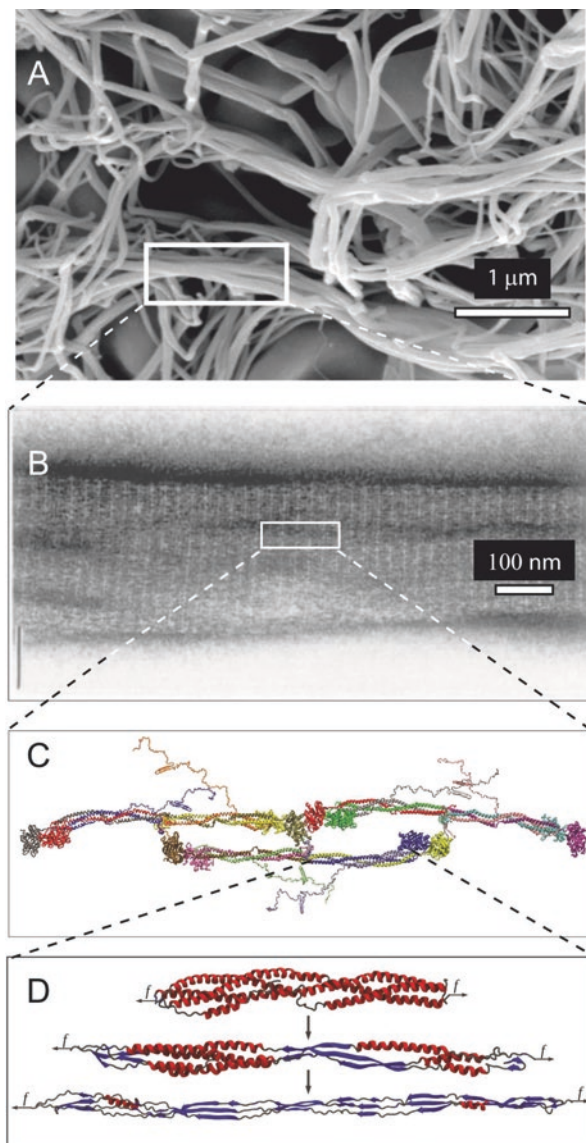
recently. The secondary structure changes during deformations of fibrin polymers studied using Fourier Transform infrared spectroscopy showed that both extension and compression of a hydrated fibrin clot are accompanied by the transition of a triple α-helix to β-sheet (Litvinov et al. 2012). This structural transition has been confirmed and analyzed in detail using full-atom molecular dynamics simulations of extensional mechanical unfolding of a fibrin(ogen) molecule (Zhmurov et al. 2012; Fig. 13.10).

Based on the paracrystalline molecular packing within a fibrin fiber that results in a 22.5-nm periodicity, it has been hypothesized that forced elongation and unfolding of fibrin molecules during stretching must result in an increase of this repeat. However, small angle X-ray scattering revealed that the position of the peak corresponding to the 22.5-nm spacing does not change significantly as the clot is stretched. Instead, there was an increase in sample disorder, consistent with an increasing number of molecules unfolding non-uniformly in response to the large strain, as expected for a two-state system in which some molecules extend completely while others remain folded (Brown et al. 2009).

Single-stranded fibrinogen oligomers have been prepared and the unfolding of fibrinogen domains has been measured by single-molecule atomic force microscopy (Brown et al. 2007; Zhmurov et al. 2011). In addition, computer-based modeling of the experimental data enabled the visualization of the various structural transitions and identification of relationships to the forces observed in the experimental or simulated force-extension curves. Molecular elongation of fibrin(ogen) is largely determined by the combined sequential unfolding transitions in the C-terminal γ chain nodules and limited reversible extension-contraction of the α-helical coiled-coil connectors. The coiled-coils act as molecular springs to take up the slack as other domains unfold.

The role of the αC polymers in fibrin mechanics, underestimated in the past, has been studied intensively more recently (Houser et al. 2010; Falvo et al. 2010; Ping et al. 2011; Helms et al. 2012; Averett et al. 2012). It has been shown that the trans-glutaminase-catalyzed covalent crosslinking of the α chains contributes substantially to the fibrin clot stiffness and elasticity (Collet et al. 1999; Piechocka et al. 2010; Helms et al. 2012; Duval et al. 2014; Kurniawan et al. 2014), including the fast elastic recoil of stretched fibrin fibers (Hudson et al. 2013). Based on the exceptional role of the αC regions in fibrin mechanics, a structural model of fibrinogen has been proposed that captures the stress-strain behavior of individual fibrin fibers (Averett et al. 2012).

13.5.6 Fibrin as a Biomaterial

A new impetus to study fibrin mechanics came from its biomedical applications as a hydrogel (Janmey et al. 2009). Fibrin has been extensively used as a biopolymer scaffold in tissue engineering, taking advantage of its unique biological and physical characteristics, including its porosity, deformability, elasticity and biodegradability (Ahmed et al. 2008). Fibrin alone, or in combination with other materials, has been used as a biological scaffold for stem or primary cells to regenerate adipose tissue, bone, cardiac tissue, cartilage, liver, nervous tissue, ocular tissue, skin, tendons, and ligaments. Although fibrin can be used as the only biomaterial, it has been shown that the addition of collagen to fibrin can provide extra stiffness and durability of the forming biomaterial (Lai et al. 2012). Gels based on fibrin, alone or in combination with gelatin, collagen, or elastin, have been used to optimize cellular

activities, including differentiation, proliferation, and changes in morphology (Shevchenko et al. 2010; Chaterji et al. 2007; Dikovsky et al. 2006; Natesan et al. 2011; Kuehn et al. 2014; Brougham et al. 2015). The ability of cells to modify their phenotype and behavior in response to variations in the stiffness of a fibrin-containing substrate (called mechanosensing) has been used, e.g., in regulation of neurite outgrowth in fibrin gels (Man et al. 2011) and modulation of the secretory activity of endothelial cells and mesenchymal stem cells (Rao et al. 2012). Fibrin clots with longitudinally aligned fibrin fibers could serve as a scaffold for longitudinal axonal regrowth in cases of traumatic peripheral nerve injuries (Gessmann et al. 2016).

Fibrin sealants formed by mixing fibrinogen and thrombin are widely used to stop bleeding (Corral et al. 2016). Importantly, various commercially available fibrin sealants have diverse mechanical properties (Hickerson et al. 2011) and glues with distinct structure and stiffness had different hemostatic efficacy (Fortelny et al. 2011).

13.6 Lytic Stability of Fibrin

13.6.1 Molecular Mechanisms of Fibrinolysis

After a clot has formed *in vivo* and fulfilled its hemostatic function, it is normally dissolved by the fibrinolytic system, to restore the impaired blood flow (Weisel and Litvinov 2008, 2014; Longstaff and Kolev 2015). The main enzyme involved is plasmin (Pn), a serine protease derived from its inactive precursor, plasminogen (Plg), through the action of activators. Pn also cleaves several other substrates, including extracellular matrix proteins, and activates some proteases and growth factors. So, in addition to fibrinolysis, Plg and Pn are involved in several other physiological and pathophysiological processes, such as wound healing, inflammation, cell migration, angiogenesis, embryogenesis, ovulation, tumor growth and metastasis, and atherosclerosis.

There is a complex system of biochemical reactions comprising fibrinolysis and its regulatory mechanisms (Fig. 13.11). Fibrinolysis occurs first by the conversion of Plg to Pn by a Plg activator, primarily on the fibrin surface, and then by the degradation of fibrin by Pn. Any Pn that dissociates from fibrin is rapidly inactivated by inhibitors. The activators of Plg are the serine proteases tissue-type plasminogen activator (t-PA) or urokinase-type plasminogen activator and bacterial proteins that acquire proteolytic activity after the interaction with human Plg or Pn, streptokinase and staphylokinase. t-PA and staphylokinase are fibrin-selective, remaining bound to fibrin and protected from rapid inhibition, while streptokinase and two-chain urokinase-type plasminogen activator are non-fibrin-selective enzymes, activating both Plg in the circulating blood and fibrin-bound Plg. Attachment of Plg and t-PA to fibrin is mediated by the C-terminal lysine residues of fibrin and the specific lysine-binding motifs on the Plg and t-PA molecules. Because Pn cleaves at lysine
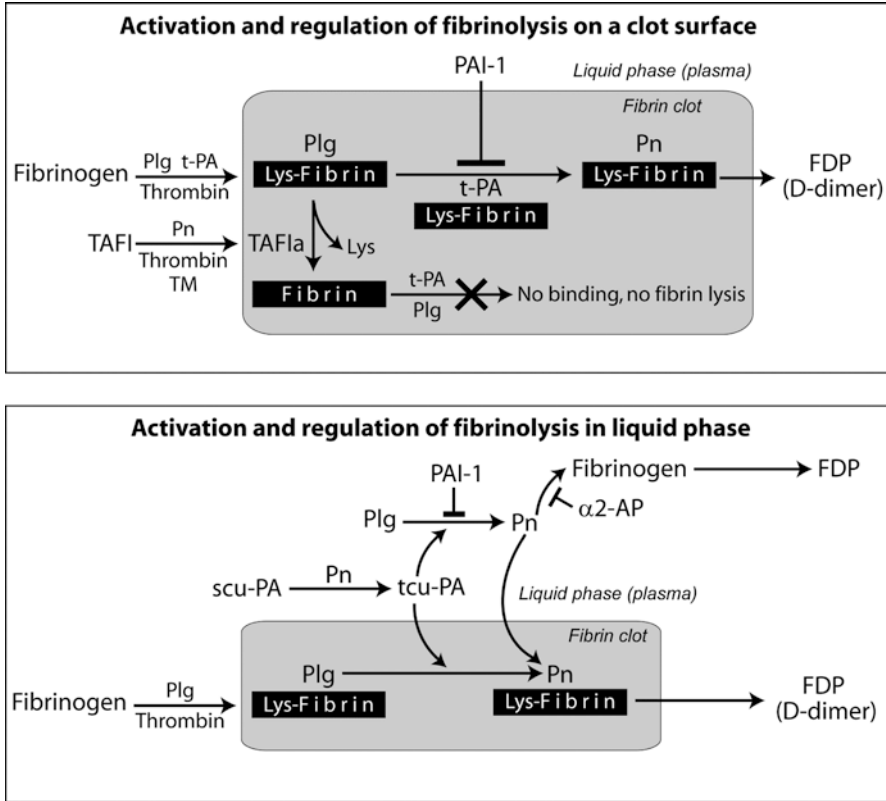
Fig. 13.11 Schematic diagram of fibrinolysis on the fibrin clot surface and in the liquid phase. Schematic representation of the major reactions of fibrinolysis and their regulation on a fibrin clot surface and in the surrounding plasma milieu. The *grey* highlighted area represents a fibrin clot surface (solid phase) surrounded by the blood plasma (liquid phase). *Black arrows* show the biochemical conversions involving proteolytic cleavage. *T*-like symbols indicate inhibitory effects. Abbreviations and functions of molecules: *Lys-Fibrin* C-terminal lysine residues on fibrin to which Plg and t-PA bind selectively, *Plg* plasminogen, bound to the C-terminal lysine residues on fibrin and free in plasma, *Pn* plasmin, formed on fibrin (by the action of t-PA) and in plasma (by the action of tcu-PA) from Plg. Pn cleaves fibrin and fibrinogen, activates scu-PA and TAFI, *t-PA* tissue-type Plg activator, fibrin-selective Plg activator, bound to fibrin via the C-terminal lysine residues on fibrin, *scu-PA* single-chain urokinase-type Plg activator (inactive), *tcu-PA* two-chain u-PA (active), non-fibrin-selective Plg activator, *PAI-1* plasminogen activator inhibitor-1, blocks both t-PA and tcu-PA, *TAFIa* thrombin-activatable fibrinolysis inhibitor (enzymatically active form) that splits off the C-terminal lysine residues from fibrin, thus preventing binding of Plg and t-PA to fibrin, α_2-*AP* α_2-antiplasmin, direct Pn inhibitor, forms circulating Pn-α_2-AP complexes, *FDP* fibrin(ogen) degradation products, resulting from cleavage of fibrin or fibrinogen by Pn, *D-dimer* a proteolytic fragment (degradation product) that is formed by Pn only from crosslinked fibrin

residues, the new C-terminal lysines provide additional binding sites for Plg and t-PA, as a positive feedback mechanism. The number of t-PA and Plg binding sites on fibrin and the rate of lysis can be increased by ultrasound-induced perturbation of fibrin clots, which may be of practical importance for thrombolytic therapy (Chernysh et al. 2015).

Enzymatic lysis of fibrin results in formation of soluble degradation products. Basically, the fibrin degradation products generated by plasmin are very similar to the proteolytic fragments of fibrinogen, namely intermediate fragments X and Y and the terminal fragments D and E, with the difference based on the presence or absence of fibrinopeptides (Marder and Budzynski 1975). What makes fibrin degradation products formed *in vivo* different from those of fibrinogen is the non-covalent interactions and covalent Factor XIIIa-mediated crosslinking of fibrin, which results in formation of various soluble oligomeric structures (Veklich et al. 1998; Horan and Francis 2001). The smallest crosslinked fibrin degradation product is D-dimer formed of two D regions of the adjacent fibrin molecules connected by a γ-γ bond. Because the only source of D-dimer is crosslinked fibrin, a high level of D-dimer-containing products in blood is widely used by clinicians as a laboratory sign of intravascular fibrin deposition occurring during local thrombosis or disseminated intravascular coagulation (Bates 2012). Fibrin degradation products comprise fibrin(ogen) fragments with incomplete sets of binding sites that competitively inhibit fibrin polymerization and can result in formation of large oligomeric structures containing non-polymerized fibrin molecules named "soluble fibrin complexes" or "soluble fibrin".

13.6.2   Modulators of Fibrinolysis

There are a number of biochemical reactions that moderate the activity of the pro-fibrinolytic components, such as Plg activator inhibitor-1, a potent inhibitor of t-PA and u-PA, and α_2-antiplasmin, which directly inhibits Pn (Fig. 13.11). Thrombin-activatable fibrinolysis inhibitor cleaves C-terminal lysine residues from partially degraded fibrin and thus inhibits fibrinolysis by preventing the lysine-dependent binding of Plg to fibrin. Lipoprotein(a) has a structural homology with Plg and, therefore can compete with Plg for binding to lysine residues and impair fibrinolysis (Angles-Cano et al. 2001).

The effectiveness of fibrinolysis results from the combination of regulated enzymatic activity and the physical characteristics of the fibrin scaffold, such as the density of fibers and branch points, pore size, and fiber diameter. In general, the rate of lysis appears to be faster for clots made up of thicker fibers than for clots made up of thinner fibers, but also depends on other biophysical properties of the clot (Weisel 2007). In addition, platelet aggregation and clot retraction have dramatic effects on fibrinolysis (Collet et al. 2002). Stretching also affects the rate of lysis of clots (Varju et al. 2011).

13.6.3   Internal and External Fibrinolysis

A variety of systems have been used to study fibrinolysis experimentally, and each has strengths and weaknesses and is suited to the investigation of different aspects of lysis. With each of these systems, there are often several biochemical or structural methods that can be used to quantify rates of lysis.

The process that mimics the physiological process of fibrinolysis has been called internal or intrinsic lysis. To simulate this process *in vitro*, either t-PA mixed with plasma or t-PA and Plg mixed with fibrinogen have been clotted, so that the clot is formed and then dissolved as the t-PA activates Plg to Pn on the fibrin surface. The process of lysis has been followed by measurement of the decrease of turbidity or by confocal or other light microscopy or biochemically by the appearance of lysis products (Gabriel et al. 1992; Collet et al. 2003).

The process that mimics clinical thrombolysis is called external or extrinsic fibrinolysis. For therapeutic thrombolysis, t-PA is introduced into the vasculature, so that it circulates and binds to thrombi, activating Plg to Pn on the fibrin surface. *In vitro*, t-PA can be introduced at the edge of a pre-formed clot in a chamber, commonly made of a light microscope slide and cover slip, so that lysis can be observed either by eye or by light microcopy (Sakharov and Rijken 1995; Collet et al. 2000). In the external lysis system just described, the t-PA enters the clot only by diffusion. However, in thrombolysis the t-PA is delivered by flowing blood, so there usually is permeation or perfusion into the thrombus. Thus, a variation of external lysis is to allow the t-PA to permeate into the clot. The results can be quantified by the measurement of digestion products released or by observation of changes in clot structure as a function of time.

13.7   Conclusions

During the past 10 years, all of these different facets of research have contributed to the vigorous flowering of new information about fibrinogen and fibrin. It is now well established that fibrin is essential for hemostasis and is a major factor in thrombosis, inflammation and infection. Many of the molecular mechanisms of the biological processes of hemostasis and how they go awry in thrombosis have been determined. The binding of fibrinogen to the integrin αIIbβ3 and its role in platelet aggregation continues to represent a model system to study cell adhesion. In addition, fibrin has become increasingly used as a versatile and unique biomaterial.

In spite of the breadth and depth of research on fibrin(ogen) in the past 10 years, many mysteries remain. For example, molecular mechanisms of lateral aggregation and branching in fibrin polymerization are largely unknown. Since the preponderance of research on fibrinogen and fibrin has been on mammalian systems, much remains unknown about their roles in other animals. Many of the binding partners of fibrin(ogen) have been identified but their functions are mostly unknown. Basic

mechanisms of the catabolism of fibrinogen remain to be discovered. The burgeoning field of fibrin mechanics will continue to thrive. Much remains to be learned about biological roles of fibrinogen beyond hemostasis. In the future, research will increasingly investigate major aspects of the role of fibrin(ogen) in more complex biological processes, such as wound healing and inflammation. The diverse array of powerful genetic and synthetic tools in various model systems, together with new super-resolution, single-molecule and single-cell imaging and manipulation methods, will drive this research. One thing is certain, fibrinogen and fibrin will continue to surprise us in the future, as they have in the past.

References

Ahmed TE, Dare EV, Hincke M (2008) Fibrin: a versatile scaffold for tissue engineering applications. Tissue Eng Part B Rev 14:199–215

Ajjan R, Lim BC, Standeven KF, Harrand R, Dolling S, Phoenix F, Greaves R, Abou-Saleh RH, Connell S, Smith DA, Weisel JW, Grant PJ, Ariens RA (2008) Common variation in the C-terminal region of the fibrinogen beta-chain: effects on fibrin structure, fibrinolysis and clot rigidity. Blood 111:643–650

Ajjan RA, Standeven KF, Khanbhai M, Phoenix F, Gersh KC, Weisel JW, Kearney MT, Ariens RA, Grant PJ (2009) Effects of aspirin on clot structure and fibrinolysis using a novel in vitro cellular system. Arterioscler Thromb Vasc Biol 29:712–717

Aleman MM, Walton BL, Byrnes JR, Wolberg AS (2014) Fibrinogen and red blood cells in venous thrombosis. Thromb Res 133(Suppl 1):S38–S40

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 10:1072–1080

Alves CS, Yakovlev S, Medved L, Konstantopoulos K (2009) Biomolecular characterization of CD44-fibrin(ogen) binding: distinct molecular requirements mediate binding of standard and variant isoforms of CD44 to immobilized fibrin(ogen). J Biol Chem 284:1177–1189

Angles-Cano E, De La Pena DA, Loyau S (2001) Inhibition of fibrinolysis by lipoprotein(a). Ann N Y Acad Sci 936:261–275

Ariens RA (2013) Fibrin(ogen) and thrombotic disease. J Thromb Haemost 11(Suppl 1):294–305

Ariens RA, Lai T-S, Weisel JW, Greenberg CS, Grant PJ (2002) Role of Factor XIII in fibrin clot formation and effects of genetic polymorphisms. Blood 100:743–754

Ashton JH, Vande Geest JP, Simon BR, Haskett DG (2009) Compressive mechanical properties of the intraluminal thrombus in abdominal aortic aneurysms and fibrin-based thrombus mimics. J Biomech 42:197–201

Asselta R, Duga S, Tenchini ML (2006) The molecular basis of quantitative fibrinogen disorders. J Thromb Haemost 4:2115–2129

Asselta R, Plate M, Robusto M, Borhany M, Guella I, Solda G, Afrasiabi A, Menegatti M, Shamsi T, Peyvandi F, Duga S (2015) Clinical and molecular characterisation of 21 patients affected by quantitative fibrinogen deficiency. Thromb Haemost 113:567–576

Averett LE, Geer CB, Fuierer RR, Akhremitchev BB, Gorkun OV, Schoenfisch MH (2008) Complexity of "A-a" knob-hole fibrin interaction revealed by atomic force spectroscopy. Langmuir 24:4979–4988

Averett LE, Schoenfisch MH, Akhremitchev BB, Gorkun OV (2009) Kinetics of the multistep rupture of fibrin 'A-a' polymerization interactions measured using atomic force microscopy. Biophys J 97:2820–2828

Averett RD, Menn B, Lee EH, Helms CC, Barker T, Guthold M (2012) A modular fibrinogen model that captures the stress-strain behavior of fibrin fibers. Biophys J 103:1537–1544

Bacon-Baguley T, Ogilvie ML, Gartner TK, Walz DA (1990) Thrombospondin binding to specific sequences within the A alpha- and B beta-chains of fibrinogen. J Biol Chem 265:2317–2323

Bailey K, Astbury WT, Rudall KM (1943) Fibrinogen and fibrin as members of the keratin-myosin group. Nature 151:716–717

Bates SM (2012) D-dimer assays in diagnosis and management of thrombotic and bleeding disorders. Semin Thromb Hemost 38:673–682

Bennett JS (2001) Platelet-fibrinogen interactions. Ann N Y Acad Sci 936:340–354

Blombäck B, Okada M (1982) Fibrin gel structure and clotting time. Thromb Res 25:51–70

Blombäck B, Hessel B, Hogg D, Therkildsen L (1978) A two-step fibrinogen-fibrin transition in blood coagulation. Nature 275:501–505

Bowley SR, Okumura N, Lord ST (2009) Impaired protofibril formation in fibrinogen gamma N308 K is due to altered D:D and "A:a" interactions. Biochemistry 48:8656–8663

Brass LF, Diamond SL (2016) Transport physics and biorheology in the setting of haemostasis and thrombosis. J Thromb Haemost 14(5):906–917

Brennan SO (2015) Variation of fibrinogen oligosaccharide structure in the acute phase response: possible haemorrhagic implications. BBA Clin 3:221–226

Brennan SO, Fellowes AP, George PM (2001) Molecular mechanisms of hypo- and afibrinogenemia. Ann N Y Acad Sci 936:91–100

Brennan SO, Davis RL, Mosesson MW, Hernandez I, Lowen R, Alexander SJ (2007) Congenital hypodysfibrinogenaemia (Fibrinogen Des Moines) due to a gamma320Asp deletion at the Ca2+ binding site. Thromb Haemost 98:467–469

Bridge KI, Philippou H, Ariens R (2014) Clot properties and cardiovascular disease. Thromb Haemost 112:901–908

Brougham CM, Levingstone TJ, Jockenhoevel S, Flanagan TC, O'brien FJ (2015) Incorporation of fibrin into a collagen-glycosaminoglycan matrix results in a scaffold with improved mechanical properties and enhanced capacity to resist cell-mediated contraction. Acta Biomater 26:205–214

Brown JH, Volkmann N, Jun G, Henschen-Edman AH, Cohen C (2000) The crystal structure of modified bovine fibrinogen. Proc Natl Acad Sci U S A 97:85–90

Brown AE, Litvinov RI, Discher DE, Weisel JW (2007) Forced unfolding of coiled-coils in fibrinogen by single-molecule AFM. Biophys J 92:L39–L41

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 325:741–744

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

Bucay I, O'brien ET, Wulfe SD, Superfine R, Wolberg AS, Falvo MR, Hudson NE (2015) Physical determinants of fibrinolysis in single fibrin fibers. PLoS One 10:e0116350

Campbell RA, Aleman M, Gray LD, Falvo MR, Wolberg AS (2010) Flow profoundly influences fibrin network structure: implications for fibrin formation and clot stability in haemostasis. Thromb Haemost 104:1281–1284

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 89:632–636

Carlisle CR, Sparks EA, Der Loughian C, Guthold M (2010) Strength and failure of fibrin fiber branchpoints. J Thromb Haemost 8:1135–1138

Casini A, Neerman-Arbez M, Ariens RA, De Moerloose P (2015) Dysfibrinogenemia: from molecular anomalies to clinical manifestations and management. J Thromb Haemost 13:909–919

Casini A, Duval C, Pan X, Tintillier V, Biron-Andreani C, Ariens RA (2016) Fibrin clot structure in patients with congenital dysfibrinogenaemia. Thromb Res 137:189–195

Chan LW, Wang X, Wei H, Pozzo LD, White NJ, Pun SH (2015) A synthetic fibrin cross-linking polymer for modulating clot properties and inducing hemostasis. Sci Transl Med 7:277ra229

Chaterji S, Kwon IK, Park K (2007) Smart polymeric gels: redefining the limits of biomedical devices. Prog Polym Sci 32:1083–1122

Chernysh IN, Weisel JW (2008) Dynamic imaging of fibrin network formation correlated with other measures of polymerization. Blood 111:4854–4861

Chernysh IN, Nagaswami C, Weisel JW (2011) Visualization and identification of the structures formed during early stages of fibrin polymerization. Blood 117:4609–4614

Chernysh IN, Nagaswami C, Purohit PK, Weisel JW (2012) Fibrin clots are equilibrium polymers that can be remodeled without proteolytic digestion. Sci Rep 2:879

Chernysh IN, Everbach CE, Purohit PK, Weisel JW (2015) Molecular mechanisms of the effect of ultrasound on the fibrinolysis of clots. J Thromb Haemost 13:601–609

Chung DW, Davie EW (1984) Gamma and gamma' chains of human fibrinogen are produced by alternative RNA splicing. Biochemist 23:4232–4236

Chung DW, Rixon MW, Que BG, Davie EW (1983) Cloning of fibrinogen genes and their cDNA. Ann N Y Acad Sci 408:449–456

Chung DW, Harris JE, Davie EW (1990) Nucleotide sequences of the three genes coding for human fibrinogen. Adv Exp Med Biol 281:39–48

Cilia La Corte AL, Philippou H, Ariens RA (2011) Role of fibrin structure in thrombosis and vascular disease. Adv Protein Chem Struct Biol 83:75–127

Cines DB, Lebedeva T, Nagaswami C, Hayes V, Massefski W, Litvinov RI, Rauova L, Lowery TJ, Weisel JW (2014) Clot contraction: compression of erythrocytes into tightly packed polyhedra and redistribution of platelets and fibrin. Blood 123:1596–1603

Cohen C, Parry DAD (1990) Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. Proteins 7:1–15

Colace TV, Muthard RW, Diamond SL (2012) Thrombus growth and embolism on tissue factor-bearing collagen surfaces under flow: role of thrombin with and without fibrin. Arterioscler Thromb Vasc Biol 32:1466–1476

Collen D, Tytgat GN, Claeys H, Piessens R (1972) Metabolism and distribution of fibrinogen. I. Fibrinogen turnover in physiological conditions in humans. Br J Haematol 22:681–700

Coller BS (2011) Historical perspective and future directions in platelet research. J Thromb Haemost 9(Suppl 1):374–395

Coller BS, Shattil SJ (2008) The GPIIb/IIIa (integrin alphaIIbbeta3) odyssey: a technology-driven saga of a receptor with twists, turns, and even a bend. Blood 112:3011–3025

Collet J-P, Veklich Y, Mullin JL, Gorkun OV, Lord ST, Weisel JW (1999) The αC domains of fibrinogen affect the structure of the clot and its physical and biochemical properties. Thromb Haemost 82:Suppl. 692

Collet J-P, Park D, Lesty C, Soria J, Soria C, Montalescot G, Weisel JW (2000) Influence of fibrin network conformation and fibrin fiber diameter on fibrinolysis speed: dynamic and structural approaches by confocal microscopy. Arterioscler Thromb Vasc Biol 20:1354–1361

Collet J-P, Montalescot G, Lesty C, Weisel JW (2002) A structural and dynamic investigation of the facilitating effect of glycoportein IIb/IIIa inhibitors in dissolving platelet-rich clots. Circ Res 90:428–434

Collet J-P, Lesty C, Montalescot G, Weisel JW (2003) Dynamic changes of fibrin architecture during fibrin formation and intrinsic fibrinolysis of fibrin-rich clots. J Biol Chem 278:21331–21335

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 106:3824–3830

Collet J-P, Allali Y, Lesty C, Tanguy ML, Silvain J, Ankri A, Blanchet B, Dumaine R, Giannetti J, Payot L, Weisel JW, Montalescot G (2006) Altered fibrin architecture is associated with hypo-fibrinolysis and premature coronary artery atherothrombosis. Arterioscler Thromb Vasc Biol 26:2567–2573

Cooper AV, Standeven KF, Ariens RA (2003) Fibrinogen gamma-chain splice variant gamma' alters fibrin formation and structure. Blood 102:535–540

Corral M, Ferko N, Hollmann S, Hogan A, Jamous N, Batiller J, Shen J (2016) Clinician reported ease of use for a novel fibrin sealant patch for hemostasis: results from four randomized controlled trials. Curr Med Res Opin 32:367–375

Cote HC, Lord ST, Pratt KP (1998) Gamma-chain dysfibrinogenemias: molecular structure-function relationships of naturally occurring mutations in the gamma chain of human fibrinogen. Blood 92:2195–2212

Crabtree GR (1987) The molecular biology of fibrinogen. In: Stamatoyannopoulos G, Nienhuis AW, Leder P, Majerus PE (eds) The molecular basis of blood diseases. W.B. Saunders, Philadelphia

Dang CV, Shin CK, Bell WR, Nagaswami C, Weisel JW (1989) Fibrinogen sialic acid residues are low affinity calcium-binding sites that influence fibrin assembly. J Biol Chem 264:15104–15108

Davies NA, Harrison NK, Morris RH, Noble S, Lawrence MJ, D'silva LA, Broome L, Brown MR, Hawkins KM, Williams PR, Davidson S, Evans PA (2015) Fractal dimension (df) as a new structural biomarker of clot microstructure in different stages of lung cancer. Thromb Haemost 114:1251–1259

De Maat MP, Verschuur M (2005) Fibrinogen heterogeneity: inherited and noninherited. Curr Opin Hematol 12:377–383

De Moerloose P, Neerman-Arbez M (2009) Congenital fibrinogen disorders. Semin Thromb Hemost 35:356–366

De Moerloose P, Casini A, Neerman-Arbez M (2013) Congenital fibrinogen disorders: an update. Semin Thromb Hemost 39:585–595

Dikovsky D, Bianco-Peled H, Seliktar D (2006) The effect of structural alterations of PEG-fibrinogen hydrogel scaffolds on 3-D cellular morphology and cellular migration. Biomaterials 27:1496–1506

Domingues MM, Macrae FL, Duval C, Mcpherson HR, Bridge KI, Ajjan RA, Ridger VC, Connell SD, Philippou H, Ariens RA (2015) Thrombin and fibrinogen gamma' impact clot structure by marked effects on intrafibrillar structure and protofibril packing. Blood 127(4):487–495

Donovan JW, Mihalyi E (1985) Clotting of fibrinogen. 1. Scanning calorimetric study of the effect of calcium. Biochemist 24:3434–3443

Doolittle RF (1984) Fibrinogen and fibrin. Annu Rev Biochem 53:195–229

Doolittle LR, 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. Biochemist 45:2657–2667

Dunn EJ, Ariens RA (2004) Fibrinogen and fibrin clot structure in diabetes. Herz 29:470–479

Duval C, Allan P, Connell SD, Ridger VC, Philippou H, Ariens RA (2014) Roles of fibrin alpha- and gamma-chain specific cross-linking by FXIIIa in fibrin structure and function. Thromb Haemost 111:842–850

Dyr JE, Blombäck B, Hessel B, Kornalik F (1989) Conversion of fibrinogen to fibrin induced by preferential release of fibrinopeptide B. Biochim Biophys Acta 990:18–24

Erickson HP, Fowler WE (1983) Electron microscopy of fibrinogen, its plasmic fragments and small polymers. Ann N Y Acad Sci 408:146–163

Evans PA, Hawkins K, Morris RH, Thirumalai N, Munro R, Wakeman L, Lawrence MJ, Williams PR (2010) Gel point and fractal microstructure of incipient blood clots are significant new markers of hemostasis for healthy and anticoagulated blood. Blood 116:3341–3346

Everse SJ, Spraggon G, Doolittle RF (1998a) A three-dimensional consideration of variant human fibrinogens. Thromb Haemost 80:1–9

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. Biochemist 37:8637–8642

Everse SJ, Spraggon G, Veerapandian L, Doolittle RF (1999) Conformational changes in fragments D and double-D from human fibrin(ogen) upon binding the peptide ligand Gly-His-Arg-Pro-amide. Biochemist 38:2941–2946

Falvo MR, Gorkun OV, Lord ST (2010) The molecular origins of the mechanical properties of fibrin. Biophys Chem 152:15–20

Ferri F, Greco M, Arcovito G, De Spirito M, Rocco M (2002) Structure of fibrin gels studied by elastic light scattering techniques: dependence of fractal dimension, gel crossover length, fiber diameter, and fiber density on monomer concentration. Phys Rev E Stat Nonlin Soft Matter Phys 66(1 Pt 1):011913

Fish RJ, Neerman-Arbez M (2012) Fibrinogen gene regulation. Thromb Haemost 108:419–426

Flamm MH, Diamond SL (2012) Multiscale systems biology and physics of thrombosis under flow. Ann Biomed Eng 40:2355–2364

Fogelson AL, Keener JP (2010) Toward an understanding of fibrin branching structure. Phys Rev E Stat Nonlinear Soft Matter Phys 81:051922

Fortelny RH, Petter-Puchner AH, Ferguson J, Gruber-Blum S, Brand J, Mika K, Redl H (2011) A comparative biomechanical evaluation of hernia mesh fixation by fibrin sealant. J Surg Res 171:576–581

Fowler WE, Erickson HP (1979) Trinodular structure of fibrinogen: confirmation by both shadowing and negative-stain electron microscopy. J Mol Biol 134:241–249

Francis CW, Marder VJ, Martin SE (1980) Demonstration of a large molecular weight variant of the gamma chain of normal human plasma fibrinogen. J Biol Chem 255:5599–5604

Fu Y, Grieninger G (1994) Fib420: a normal human variant of fibrinogen with two extended alpha chains. Proc Natl Acad Sci U S A 91:2625–2628

Gabriel DA, Muga K, Boothroyd EM (1992) The effect of fibrin structure on fibrinolysis. J Biol Chem 267:24259–24263

Gailit J, Ruoslahti E (1988) Regulation of the fibronectin receptor affinity by divalent cations. J Biol Chem 263:12927–12932

Galanakis DK (1993) Inherited dysfibrinogenemia: emerging abnormal structure associations with pathologic and nonpathologic dysfunctions. Semin Thromb Hemost 19:386–395

Galanakis DK, Lane BP, Simon SR (1987) Albumin modulates lateral assembly of fibrin polymers: evidence of enhanced fine fibril formation and of unique synergism with fibrinogen. Biochemist 26:2389–2400

Galanakis DK, Henschen A, Peerschke EI, Kehl M (1989) Fibrinogen Stony Brook, a heterozygous Aalpha16Arg —> Cys dysfibrinogenemia – evaluation of diminshed platelet aggregation support and of enhanced inhibition of fibrin assembly. J Clin Invest 84:295–304

Galanakis D, Spitzer S, Scharrer I (1993) Unusual A alpha 16Arg-->Cys dysfibrinogenaemic family: absence of normal Aalpha-chains in fibrinogen from two of four heterozygous siblings. Blood Coagul Fibrinolysis 4:67–71

Galanakis DK, Nuovo G, Spitzer S, Kaplan C, Scharrer I (1996) Fibrinogen mRNA and antigen co-present in human trophoblasts in situ: possible implications. Thromb Res 81:263–269

Gasser TC, Gorgulu G, Folkesson M, Swedenborg J (2008) Failure properties of intraluminal thrombus in abdominal aortic aneurysm under static and pulsating mechanical loads. J Vasc Surg 48:179–188

Geer CB, Tripathy A, Schoenfisch MH, Lord ST, Gorkun OV (2007) Role of 'B-b' knob-hole interactions in fibrin binding to adsorbed fibrinogen. J Thromb Haemost 5:2344–2351

Gersh KC, Nagaswami C, Weisel JW, Lord ST (2009) The presence of gamma' chain impairs fibrin polymerization. Thromb Res 124:356–363

Gersh KC, Edmondson KE, Weisel JW (2010) Flow rate and fibrin fiber alignment. J Thromb Haemost 8:2826–2828

Gessmann J, Seybold D, Peter E, Schildhauer TA, Koller M (2016) Alignment of the fibrin network within an autologous plasma clot. Tissue Eng Part C Methods 22:30–37

Gilbert GE, Novakovic VA, Shi J, Rasmussen J, Pipe SW (2015) Platelet binding sites for factor VIII in relation to fibrin and phosphatidylserine. Blood 126:1237–1244

Gorkun OV, Litvinov RI, Veklich YI, Weisel JW (2006) Interactions mediated by the N-terminus of fibrinogen's Bbeta chain. Biochemistry 45:14843–14852

Guo YH, Hernandez I, Isermann B, Kang TB, Medved L, Sood R, Kerschen EJ, Holyst T, Mosesson MW, Weiler H (2009) Caveolin-1-dependent apoptosis induced by fibrin degradation products. Blood 113:4431–4439

Guthold M, Liu W, Stephens B, Lord ST, Hantgan RR, Erie DA, Taylor RM, Superfine R (2004) Visualization and mechanical manipulations of individual fibrin fibers suggest that fiber cross section has fractal dimension 1.3. Biophys J 87:4226–4236

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 49:165–181

Haidaris PJ (1997) Induction of fibrinogen biosynthesis and secretion from cultured pulmonary epithelial cells. Blood 89:873–882

Haidaris PJ, Courtney MA (1990) Tissue-specific and ubiquitous expression of fibrinogen gamma-chain mRNA. Blood Coagul Fibrinolysis 1:433–437

Haidaris PJ, Francis CW, Sporn LA, Arvan DS, Collichio FA, Marder VJ (1989) Megakaryocyte and hepatocyte origins of human fibrinogen biosynthesis exhibit hepatocyte-specific expression of gamma chain-variant polypeptides. Blood 74:743–750

Hall CE, Slayter HS (1959) The fibrinogen molecule: its size, shape and mode of polymerization. J Biophys Biochem Cytol 5:11–16

Hantgan RR, Simpson-Haidaris PJ, Francis CW, Marder VJ (2000) Fibrinogen structure and physiology. In: RW C, Hirsh J, VJ M, AW C, JN G (eds) Hemostasis and thrombosis: basic principles and clinical practice, 4th edn. Lippincott, Williams & Wilkins, Philadelphia

Harrison P, Wilbourn B, Debili N, Vainchenker W, Breton-Gorius J, Lawrie AS, Masse JM, Savidge GF, Cramer EM (1989) Uptake of plasma fibrinogen into the alpha granules of human megakaryocytes and platelets. J Clin Invest 73:1123–1129

Harrison P, Savidge GF, Cramer EM (1990) The origin and physiological relevance of alpha-granule adhesive proteins. Br J Haematol 74:125–130

Heffron SP, Parastatidis I, Cuchel M, Wolfe ML, Tadesse MG, Mohler ER 3rd, Ischiropoulos H, Rader DJ, Reilly MP (2009) Inflammation induces fibrinogen nitration in experimental human endotoxemia. Free Radic Biol Med 47:1140–1146

Helms CC, Ariens RA, Uitte De Willige S, Standeven KF, Guthold M (2012) alpha-alpha cross-links increase fibrin fiber elasticity and stiffness. Biophys J 102:168–175

Henderson SJ, Xia J, Wu H, Stafford AR, Leslie BA, Fredenburgh JC, Weitz DA, Weitz JI (2015) Zinc promotes clot stability by accelerating clot formation and modifying fibrin structure. Thromb Haemost 115(3):533–542

Henschen A, Mcdonagh J (1986) Fibrinogen, fibrin and factor XIII. In: Zwaal RFA, Hemker HC (eds) Blood coagulation. Elsevier Science, Amsterdam

Henschen-Edman AH (2001) Fibrinogen non-inherited heterogeneity and its relationship to function in health and disease. Ann N Y Acad Sci 936:580–593

Hickerson WL, Nur I, Meidler R (2011) A comparison of the mechanical, kinetic, and biochemical properties of fibrin clots formed with two different fibrin sealants. Blood Coagul Fibrinolysis 22:19–23

Hirota-Kawadobora M, Terasawa F, Yonekawa O, Sahara N, Shimizu E, Okumura N, Katsuyama T, Shigematsu H (2003) Fibrinogens Kosai and Ogasa: Bbeta15Gly-->Cys (GGT-->TGT) substitution associated with impairment of fibrinopeptide B release and lateral aggregation. J Thromb Haemost 1:275–283

Horan JT, Francis CW (2001) Fibrin degradation products, fibrin monomer and soluble fibrin in disseminated intravascular coagulation. Semin Thromb Hemost 27:657–666

Houser JR, Hudson NE, Ping L, O'brien ET 3rd, Superfine R, Lord ST, Falvo MR (2010) Evidence that alphaC region is origin of low modulus, high extensibility, and strain stiffening in fibrin fibers. Biophys J 99:3038–3047

Howes JM, Richardson VR, Smith KA, Schroeder V, Somani R, Shore A, Hess K, Ajjan R, Pease RJ, Keen JN, Standeven KF, Carter AM (2012) Complement C3 is a novel plasma clot component with anti-fibrinolytic properties. Diab Vasc Dis Res 9:216–225

Huang L, Hsiao JP, Powierza C, Taylor RM 2nd, Lord ST (2014) Does topology drive fiber polymerization? Biochemistry 53:7824–7834

Hudson NE, Houser JR, O'brien ET 3rd, Taylor RM 2nd, Superfine R, Lord ST, Falvo MR (2010) Stiffening of individual fibrin fibers equitably distributes strain and strengthens networks. Biophys J 98:1632–1640

Hudson NE, Ding F, Bucay I, O'brien ET 3rd, Gorkun OV, Superfine R, Lord ST, Dokholyan NV, Falvo MR (2013) Submillisecond elastic recoil reveals molecular origins of fibrin fiber mechanics. Biophys J 104:2671–2680

Hudson NE, Houser JR, O'brien ET 3rd, Taylor RM 2nd, Superfine R, Lord ST, Falvo MR (2015) Stiffening of individual fibrin fibers equitably distributes strain and strengthens networks. Biophys J 98:1632–1640

Iino M, Takeya H, Takemitsu T, Nakagaki T, Gabazza EC, Suzuki K (1995) Characterization of the binding of factor Xa to fibrinogen/fibrin derivatives and localization of the factor Xa binding site on fibrinogen. Eur J Biochem 232:90–97

Jandrot-Perrus M, Mosesson MW, Denninger MH, Menache D (1979) Studies of platelet fibrinogen from a subject with a congenital plasma fibrinogen abnormality (fibrinogen Paris I). Blood 54:1109–1116

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 27:135–153

Janmey PA, Winer JP, Weisel JW (2009) Fibrin gels and their clinical and bioengineering applications. J R Soc Interface 6:1–10

Jansen KA, Bacabac RG, Piechocka IK, Koenderink GH (2013) Cells actively stiffen fibrin networks by generating contractile stress. Biophys J 105:2240–2251

Kant JA, Fornace AJ, Saxe D, Simon MI, Mcbride OW, Crabtree GR (1985) Evolution and organization of the fibrinogen locus on chromosome 4: gene duplication accompanied by transposition and inversion. Proc Natl Acad Sci U S A 185:1–19

Kim OV, Litvinov RI, Weisel JW, Alber MS (2014) Structural basis for the nonlinear mechanics of fibrin networks under compression. Biomaterials 35:6739–6749

Kohler S, Schmid F, Settanni G (2015) The internal dynamics of fibrinogen and its implications for coagulation and adsorption. PLoS Comput Biol 11:e1004346

Kolev K, Tenekedjiev K, Ajtai K, Kovalszky I, Gombas J, Varadi B, Machovich R (2003) Myosin: a noncovalent stabilizer of fibrin in the process of clot dissolution. Blood 101:4380–4386

Kononova O, Litvinov RI, Zhmurov A, Alekseenko A, Cheng CH, Agarwal S, Marx KA, Weisel JW, Barsegov V (2013) Molecular mechanisms, thermodynamics, and dissociation kinetics of knob-hole interactions in fibrin. J Biol Chem 288:22681–22692

Koopman J, Haverkate F, Lord ST, Grimbergen J, Mannucci PM (1992) Molecular basis of fibrinogen Naples associated with defective thrombin binding and thrombophilia. Homozygous substitution of B beta 68 Ala —> Thr. J Clin Invest 90:238–244

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. Biochemist 41:12124–12132

Kostelansky MS, Bolliger-Stucki B, Betts L, Gorkun OV, Lord ST (2004a) B beta Glu397 and B beta Asp398 but not B beta Asp432 are required for "B:b" interactions. Biochemist 43:2465–2474

Kostelansky MS, Lounes KC, Ping LF, Dickerson SK, Gorkun OV, Lord ST (2004b) Calcium-binding site beta 2, adjacent to the "b" polymerization site, modulates lateral aggregation of protofibrils during fibrin polymerization. Biochemist 43:2475–2483

Kostelansky MS, Lounes KC, Ping LF, Dickerson SK, Gorkun OV, Lord ST (2007) Probing the gamma2 calcium-binding site: studies with gammaD298,301A fibrinogen reveal changes in the gamma294-301 loop that alter the integrity of the "a" polymerization site. Biochemist 46:5114–5123

Kotlarchyk MA, Shreim SG, Alvarez-Elizondo MB, Estrada LC, Singh R, Valdevit L, Kniazeva E, Gratton E, Putnam AJ, Botvinick EL (2011) Concentration independent modulation of local micromechanics in a fibrin gel. PLoS One 6:e20201

Kuehn C, Fulop T, Lakey JR, Vermette P (2014) Young porcine endocrine pancreatic islets cultured in fibrin and alginate gels show improved resistance towards human monocytes. Pathol Biol (Paris) 62:354–364

Kurniawan NA, Grimbergen J, Koopman J, Koenderink GH (2014) Factor XIII stiffens fibrin clots by causing fiber compaction. J Thromb Haemost 12:1687–1696

Lai VK, Lake SP, Frey CR, Tranquillo RT, Barocas VH (2012) Mechanical behavior of collagen-fibrin co-gels reflects transition from series to parallel interactions with increasing collagen content. J Biomech Eng Trans ASME 134

Lam WA, Chaudhuri O, Crow A, Webster KD, Li TD, Kita A, Huang J, Fletcher DA (2011) Mechanics and contraction dynamics of single platelets and implications for clot stiffening. Nat Mater 10:61–66

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 263:15056–15063

Lauricella AM, Castanon MM, Kordich LC, Quintana IL (2013) Alterations of fibrin network structure mediated by dermatan sulfate. J Thromb Thrombolysis 35:257–263

Lawrence SO, Simpson-Haidaris PJ (2004) Regulated de novo biosynthesis of fibrinogen in extra-hepatic epithelial cells in response to inflammation. Thromb Haemost 92:234–243

Lawrence MJ, Sabra A, Mills G, Pillai SG, Abdullah W, Hawkins K, Morris RH, Davidson SJ, D'silva LA, Curtis DJ, Brown MR, Weisel JW, Williams PR, Evans PA (2015) A new bio-marker quantifies differences in clot microstructure in patients with venous thromboembolism. Br J Haematol 168:571–575

Lim BC, Ariens RA, Carter AM, Weisel JW, Grant PJ (2003) Genetic regulation of fibrin structure and function: complex gene-environment interactions may modulate vascular risk. Lancet 361:1424–1431

Lim BCB, Lee EH, Sotomayor M, Schulten K (2008) Molecular basis of fibrin clot elasticity. Structure 16:449–459

Lindstrom SB, Kulachenko A, Jawerth LM, Vader DA (2013) Finite-strain, finite-size mechanics of rigidly cross-linked biopolymer networks. Soft Matter 9:7302–7313

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 106:2944–2951

Litvinov RI, Gorkun OV, Galanakis DK, Yakovlev S, Medved L, Shuman H, Weisel JW (2007a) Polymerization of fibrin: direct observation and quantification of individual B:b knob-hole interactions. Blood 109:130–138

Litvinov RI, Yakovlev S, Tsurupa G, Gorkun OV, Medved L, Weisel JW (2007b) Direct evidence for specific interactions of the fibrinogen alphaC-domains with the central E region and with each other. Biochemistry 46:9133–9142

Litvinov RI, Faizullin DA, Zuev YF, Weisel JW (2012) The alpha-helix to beta-sheet transition in stretched and compressed hydrated fibrin clots. Biophys J 103:1020–1027

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 313:634

Liu W, Carlisle CR, Sparks EA, Guthold M (2010) The mechanical properties of single fibrin fibers. J Thromb Haemost 8:1030–1036

Longstaff C, Kolev K (2015) Basic mechanisms and regulation of fibrinolysis. J Thromb Haemost 13(Suppl 1):S98–S105

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 288:6946–6956

Lord ST (2007) Fibrinogen and fibrin: scaffold proteins in hemostasis. Curr Opin Hematol 14:236–241

Lord ST (2011) Molecular mechanisms affecting fibrin structure and stability. Arterioscler Thromb Vasc Biol 31:494–499

Louache F, Debili N, Cramer EM, Breton-Gorius J, Vainchenker W (1991) Fibrinogen is not synthesized by human megakaryocytes. Blood 77:311–316

Lovely RS, Falls LA, Al-Mondhiry HA, Chambers CE, Sexton GJ, Ni H, Farrell DH (2002) Association of gammaA/gamma' fibrinogen levels and coronary artery disease. Thromb Haemost 88:26–31

Ly B, Godal HC (1973) Denaturation of fibrinogen: the protective effect of calcium. Haematologica 1:204

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 U S A 98:11967–11972

Magatti D, Molteni M, Cardinali B, Rocco M, Ferri F (2013) Modeling of fibrin gels based on confocal microscopy and light-scattering data. Biophys J 104:1151–1159

Malecki R, Gacka M, Kuliszkiewicz-Janus M, Jakobsche-Policht U, Kwiatkowski J, Adamiec R, Undas A (2015) Altered plasma fibrin clot properties in essential thrombocythemia. Platelets: 1–7

Man AJ, Davis HE, Itoh A, Leach JK, Bannerman P (2011) Neurite outgrowth in fibrin gels is regulated by substrate stiffness. Tissue Eng A 17:2931–2942

Mannila MN, Eriksson P, Ericsson CG, Hamsten A, Silveira A (2006) Epistatic and pleiotropic effects of polymorphisms in the fibrinogen and coagulation factor XIII genes on plasma fibrinogen concentration, fibrin gel structure and risk of myocardial infarction. Thromb Haemost 95:420–427

Mannila MN, Lovely RS, Kazmierczak SC, Eriksson P, Samnegard A, Farrell DH, Hamsten A, Silveira A (2007) Elevated plasma fibrinogen gamma' concentration is associated with myocardial infarction: effects of variation in fibrinogen genes and environmental factors. J Thromb Haemost 5:766–773

Maquart FX, Monboisse JC (2014) Extracellular matrix and wound healing. Pathol Biol (Paris) 62:91–95

Marchi RC, Carvajal Z, Boyer-Neumann C, Angles-Cano E, Weisel JW (2006) Functional characterization of fibrinogen Bicetre II: a gamma 308 Asn-->Lys mutation located near the fibrin D:D interaction sites. Blood Coagul Fibrinolysis 17:193–201

Marder VJ, Budzynski AZ (1975) Data for defining fibrinogen in its plasmic degradation products. Thromb Diath Haemorrh 33:199–207

Marguerie G, Chagniel G, Suscillon M (1977) The binding of calcium to bovine fibrinogen. Biochim Biophys Acta 490:94–103

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 52:5491–5502

Martinez J, Keane PM, Gilman PB, Palascak JE (1983) The abnormal carbohydrate composition of the dysfibrinogenemia associated with liver disease. Ann N Y Acad Sci 408:388–396

Martinez M, Cuker A, Mills A, Lightfoot R, Fan Y, Tang WH, Hazen SL, Ischiropoulos H (2012) Nitrated fibrinogen is a biomarker of oxidative stress in venous thromboembolism. Free Radic Biol Med 53:230–236

Martinez M, Weisel JW, Ischiropoulos H (2013) Functional impact of oxidative posttranslational modifications on fibrinogen and fibrin clots. Free Radic Biol Med 65:411–418

Martinez MR, Cuker A, Mills AM, Crichlow A, Lightfoot RT, Chernysh IN, Nagaswami C, Weisel JW, Ischiropoulos H (2014) Enhanced lysis and accelerated establishment of viscoelastic properties of fibrin clots are associated with pulmonary embolism. Am J Physiol Lung Cell Mol Physiol 306:L397–L404

Matsuda M, Sugo T (2001) Hereditary disorders of fibrinogen. Ann N Y Acad Sci 936:65–68

Matsuka YV, Medved LV, Migliorini MM, Ingham KC (1996) Factor XIIIa-catalyzed cross-linking of recombinant alpha C fragments of human fibrinogen. Biochemist 35:5810–5816

Medved L, Weisel JW (2009) Recommendations for nomenclature on fibrinogen and fibrin. J Thromb Haemost 7:355–359

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 216:503–509

Medved L, Tsurupa G, Yakovlev S (2001) Conformational changes upon conversion of fibrinogen into fibrin. The mechanisms of exposure of cryptic sites. Ann N Y Acad Sci 936:185–204

Mihalyi E (1988) Clotting of bovine fibrinogen. Calcium binding to fibrin during clotting and its dependence on release of fibrinopeptide B. Biochemistry 27:967–976

Miszta A, Pelkmans L, Lindhout T, Krishnamoorthy G, De Groot PG, Hemker CH, Heemskerk JW, Kelchtermans H, De Laat B (2014) Thrombin-dependent Incorporation of von Willebrand Factor into a Fibrin Network. J Biol Chem 289:35979–35986

Moen JL, Gorkun OV, Weisel JW, Lord ST (2003) Recombinant BbetaArg14His fibrinogen implies participation of N-terminus of Bbeta chain in desA fibrin polymerization. Blood 102:2466–2471

Mosesson MW (2004) Cross-linked gamma chains in a fibrin fibril are situated transversely between its strands: yes. J Thromb Haemost 2:388–393

Mosesson MW (2007) Update on antithrombin I (fibrin). Thromb Haemost 98:105–108

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 82:1517–1521

Mosesson MW, Diorio JP, Hernandez I, Hainfeld JF, Wall JS, Grieninger G (2004) The ultrastructure of fibrinogen-420 and the fibrin-420 clot. Biophys Chem 112:209–214

Müller MF, Ris H, Ferry JD (1984) Electron microscopy of fine fibrin clots and fine and coarse fibrin films. J Mol Biol 174:369–384

Mullin JL, Gorkun OV, Lord ST (2000) Decreased lateral aggregation of a variant recombinant fibrinogen provides insight into the polymerization mechanism. Biochemistry 39:9843–9849

Munster S, Jawerth LM, Fabry B, Weitz DA (2013) Structure and mechanics of fibrin clots formed under mechanical perturbation. J Thromb Haemost 11:557–560

Muthard RW, Welsh JD, Brass LF, Diamond SL (2015) Fibrin, gamma'-fibrinogen, and transclot pressure gradient control hemostatic clot growth during human blood flow over a collagen/tissue factor wound. Arterioscler Thromb Vasc Biol 35:645–654

Natesan S, Zhang G, Baer DG, Walters TJ, Christy RJ, Suggs LJ (2011) A bilayer construct controls adipose-derived stem cell differentiation into endothelial cells and pericytes without growth factor stimulation. Tissue Eng A 17:941–953

Neerman-Arbez M (2001) Fibrinogen gene mutations accounting for congenital afibrinogenemia. Ann N Y Acad Sci 936:496–508

Neeves KB, Illing DA, Diamond SL (2010) Thrombin flux and wall shear rate regulate fibrin fiber deposition state during polymerization under flow. Biophys J 98:1344–1352

Nham SU, Fuller GM (1986) Effect of fibrinogen degradation products on production of hepatocyte stimulating factor by a macrophage cell line (P388D1). Thromb Res 44:467–475

Nossel H (1976) Radioimmunoassay of fibrinopeptides in relation to intravascular coagulation and thrombosis. NEJ Med 295:428–432

Nowak P, Zbikowska HM, Ponczek M, Kolodziejczyk J, Wachowicz B (2007) Different vulnerability of fibrinogen subunits to oxidative/nitrative modifications induced by peroxynitrite: functional consequences. Thromb Res 121:163–174

Nussenzweig V, Seligmann M, Pelmont J, Grabar P (1961) The products of degradation of human fibrinogen by plasmin. I. Separation and physicochemical properties. Ann Inst Pasteur (Paris) 100:377–389

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 105:19438–19443

Odrljin TM, Rybarczyk BJ, Francis CW, Lawrence SO, Hamaguchi M, Simpson-Haidaris PJ (1996) Calcium modulates plasmin cleavage of the fibrinogen D fragment gamma chain N-terminus: mapping of monoclonal antibody J88B to a plasmin sensitive domain of the gamma chain. Biochim Biophys Acta 1298:69–77

Okada M, Blombäck B (1983) Factors influencing fibrin gel structure studied by flow measurement. Ann N Y Acad Sci 408:233–253

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 272:29596–29601

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 Bbeta111Ser in coiled-coil region, affecting fibrin lateral aggregation. Clin Chim Acta 365:160–167

Okumura N, Terasawa F, Fujihara N, Hirota-Kawadobora M (2007) Markedly impaired but significant thrombin-catalyzed fibrin polymerization observed at variant fibrinogens at gamma364Asp residue is arisen from B-knob and b-hole bonding. J Thromb Haemost 5 (suppl 2)**:** P-W-389

Parastatidis I, Thomson L, Burke A, Chernysh I, Nagaswami C, Visser J, Stamer S, Liebler DC, Koliakos G, Heijnen HF, Fitzgerald GA, Weisel JW, Ischiropoulos H (2008) Fibrinogen beta-chain tyrosine nitration is a prothrombotic risk factor. J Biol Chem 283:33846–33853

Park CT, Wright SD (2000) Fibrinogen is a component of a novel lipoprotein particle: factor H-related protein (FHRP)-associated lipoprotein particle (FALP). Blood 95:198–204

Parrott JA, Whaley PD, Skinner MK (1993) Extrahepatic expression of fibrinogen by granulosa cells: potential role in ovulation. Endocrinology 133:1645–1649

Paton LN, Mocatta TJ, Richards AM, Winterbourn CC (2010) Increased thrombin-induced polymerization of fibrinogen associated with high protein carbonyl levels in plasma from patients post myocardial infarction. Free Radic Biol Med 48:223–229

Pechik I, Yakovlev S, Mosesson MW, Gilliland GL, Medved L (2006) Structural basis for sequential cleavage of fibrinopeptides upon fibrin assembly. Biochemist 45:3588–3597

Piechocka IK, Bacabac RG, Potters M, Mackintosh FC, Koenderink GH (2010) Structural hierarchy governs fibrin gel mechanics. Biophys J 98:2281–2289

Ping L, Huang L, Cardinali B, Profumo A, Gorkun OV, Lord ST (2011) Substitution of the human alphaC 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 50:9066–9075

Podor TJ, Campbell S, Chindemi P, Foulon DM, Farrell DH, Walton PD, Weitz JI, Peterson CB (2002) Incorporation of vitronectin into fibrin clots. Evidence for a binding interaction between vitronectin and gamma A/gamma' fibrinogen. J Biol Chem 277:7520–7528

Protopopova AD, Barinov NA, Zavyalova EG, Kopylov AM, Sergienko VI, Klinov DV (2015) Visualization of fibrinogen alphaC regions and their arrangement during fibrin network formation by high-resolution AFM. J Thromb Haemost 13:570–579

Purohit PK, Litvinov RI, Brown AE, Discher DE, Weisel JW (2011) Protein unfolding accounts for the unusual mechanical behavior of fibrin networks. Acta Biomater 7:2374–2383

Qiu Y, Brown AC, Myers DR, Sakurai Y, Mannino RG, Tran R, Ahn B, Hardy ET, Kee MF, Kumar S, Bao G, Barker TH, Lam WA (2014) Platelet mechanosensing of substrate stiffness during clot formation mediates adhesion, spreading, and activation. Proc Natl Acad Sci U S A 111:14430–14435

Rao RR, Peterson AW, Ceccarelli J, Putnam AJ, Stegemann JP (2012) Matrix composition regulates three-dimensional network formation by endothelial cells and mesenchymal stem cells in collagen/fibrin materials. Angiogenesis 15:253–264

Raynal B, Cardinali B, Grimbergen J, Profumo A, Lord ST, England P, Rocco M (2013) Hydrodynamic characterization of recombinant human fibrinogen species. Thromb Res 132:e48–e53

Redman CM, Xia H (2001) Fibrinogen biosynthesis. Assembly, intracellular degradation, and association with lipid synthesis and secretion. Ann N Y Acad Sci 936:480–495

Riedel T, Suttnar J, Brynda E, Houska M, Medved L, Dyr JE (2011) Fibrinopeptides A and B release in the process of surface fibrin formation. Blood 117:1700–1706

Roberts HR, Stinchcombe TE, Gabriel DA (2001) The dysfibringenemias. British J Haematol 114:249–257

Rocco M, Molteni M, Ponassi M, Giachi G, Frediani M, Koutsioubas A, Profumo A, Trevarin D, Cardinali B, Vachette P, Ferri F, Perez 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 136:5376–5384

Rojas AM, Kordich L, Lauricella AM (2009) Homocysteine modifies fibrin clot deformability: another possible explanation of harm. Biorheology 46:379–387

Rosenfeld MA, Shchegolikhin AN, Bychkova AV, Leonova VB, Biryukova MI, Kostanova EA (2014) Ozone-induced oxidative modification of fibrinogen: role of the D regions. Free Radic Biol Med 77:106–120

Rosenfeld MA, Leonova VB, Shchegolikhin AN, Bychkova AV, Kostanova EA, Biryukova MI (2015) Covalent structure of single-stranded fibrin oligomers cross-linked by FXIIIa. Biochem Biophys Res Commun 461:408–412

Rottenberger Z, Komorowicz E, Szabo L, Bota A, Varga Z, Machovich R, Longstaff C, Kolev K (2013) Lytic and mechanical stability of clots composed of fibrin and blood vessel wall components. J Thromb Haemost 11:529–538

Ryan EA, Mockros LF, Weisel JW, Lorand L (1999) Structural origins of fibrin clot rheology. Biophys J 77:2813–2826

Sahni A, Francis CW (2000) Vascular endothelial growth factor binds to fibrinogen and fibrin and stimulates endothelial cell proliferation. Blood 96:3772–3778

Sahni A, Sporn LA, Francis CW (1999) Potentiation of endothelial cell proliferation by fibrin(ogen)-bound fibroblast growth factor-2. J Biol Chem 274:14936–14941

Sahni A, Guo M, Sahni SK, Francis CW (2004) Interleukin-1beta but not IL-1alpha binds to fibrinogen and fibrin and has enhanced activity in the bound form. Blood 104:409–414

Sakharov DV, Rijken DC (1995) Superficial accumulation of plasminogen during plasma clot lysis. Circulation 92:1883–1890

Sauls DL, Wolberg AS, Hoffman M (2003) Elevated plasma homocysteine leads to alterations in fibrin clot structure and stability: implications for the mechanism of thrombosis in hyperhomocysteinemia. J Thromb Haemost 1:300–306

Sauls DL, Lockhart E, Warren ME, Lenkowski A, Wilhelm SE, Hoffman M (2006) Modification of fibrinogen by homocysteine thiolactone increases resistance to fibrinolysis: a potential mechanism of the thrombotic tendency in hyperhomocysteinemia. Biochemistry 45:2480–2487

Schvartz I, Seger D, Maik-Rachline G, Kreizman T, Shaltiel S (2002) Truncated vitronectins: binding to immobilized fibrin and to fibrin clots, and their subsequent interaction with cells. Biochem Biophys Res Commun 290:682–689

Scott EM, Ariens RA, Grant PJ (2004) Genetic and environmental determinants of fibrin structure and function: relevance to clinical disease. Arterioscler Thromb Vasc Biol 24:1558–1566

Shainoff JR, Dardik BN (1979) Fibrinopeptide B and aggregation of fibrinogen. Science 204:200–202

Shevchenko RV, James SL, James SE (2010) A review of tissue-engineered skin bioconstructs available for skin reconstruction. J R Soc Interface 7:229–258

Silvain J, Collet JP, Nagaswami C, Beygui F, Edmondson KE, Bellemain-Appaix A, Cayla G, Pena A, Brugier D, Barthelemy O, Montalescot G, Weisel JW (2011) Composition of coronary thrombus in acute myocardial infarction. J Am Coll Cardiol 57:1359–1367

Smolarczyk K, Boncela J, Szymanski J, Gils A, Cierniewski CS (2005) Fibrinogen contains cryptic PAI-1 binding sites that are exposed on binding to solid surfaces or limited proteolysis. Arterioscler Thromb Vasc Biol 25:2679–2684

Spero RC, Sircar RK, Schubert R, Taylor RM 2nd, Wolberg AS, Superfine R (2011) Nanoparticle diffusion measures bulk clot permeability. Biophys J 101:943–950

Spraggon G, Everse SJ, Doolittle RF (1997) Crystal structures of fragment D from human fibrinogen and its crosslinked counterpart from fibrin. Nature 389:455–462

Standeven KF, Grant PJ, Carter AM, Scheiner T, Weisel JW, Ariens RA (2003) Functional analysis of the fibrinogen Aalpha Thr312Ala polymorphism: effects on fibrin structure and function. Circulation 107:2326–2330

Standeven KF, Carter AM, Grant PJ, Weisel JW, Chernysh I, Masova L, Lord ST, Ariens RA (2007) Functional analysis of fibrin {gamma}-chain cross-linking by activated factor XIII: determination of a cross-linking pattern that maximizes clot stiffness. Blood 110:902–907

Storm C, Pastore JJ, Mackintosh FC, Lubensky TC, Janmey PA (2005) Nonlinear elasticity in biological gels. Nature 435:191–194

Sugo T, Nakamikawa C, Yoshida N, Niwa K, Sameshima M, Mimuro J, Weisel JW, Nagita A, Matsuda M (2000) End-linked homodimers in fibrinogen Osaka VI with a B beta-chain extension lead to fragile clot structure. Blood 96:3779–3785

Swieringa F, Baaten CC, Verdoold R, Mastenbroek TG, Rijnveld N, Van Der Laan KO, Breel EJ, Collins PW, Lance MD, Henskens YM, Cosemans JM, Heemskerk JW, Van Der Meijden PE

(2016) Platelet control of fibrin distribution and microelasticity in thrombus formation under flow. Arterioscler Thromb Vasc Biol 36(4):692–699. doi:10.1161/ATVBAHA.115.306537

Takahashi K, Kondo T, Yoshikawa Y, Watanabe K, Orino K (2013) The presence of heat-labile factors interfering with binding analysis of fibrinogen with ferritin in horse plasma. Acta Vet Scand 55:70

Takeda Y (1966) Studies of the metabolism and distribution of fibrinogen in healthy men with autologous 125-I-labeled fibrinogen. J Clin Invest 45:103–111

Talens S, Leebeek FW, Demmers JA, Rijken DC (2012) Identification of fibrin clot-bound plasma proteins. PLoS One 7:e41966

Tamura T, Arai S, Nagaya H, Mizuguchi J, Wada I (2013) Stepwise assembly of fibrinogen is assisted by the endoplasmic reticulum lectin-chaperone system in HepG2 cells. PLoS One 8:e74580

Torbet J, Freyssinet JM, Hudry-Clergeon G (1981) Oriented fibrin gels formed by polymerization in strong magnetic fields. Nature 289:91–93

Townsend RR, Hilliker E, Li YT, Laine RA, Bell WR, Lee YC (1982) Carbohydrate structure of human fibrinogen. Use of 300-MHz 1H-NMR to characterize glycosidase-treated glycopeptides. J Biol Chem 257:9704–9710

Townsend RR, Heller DN, Fenselau CC, Lee YC (1984) Determination of the sialylation pattern of human fibrinogen glycopeptides with fast atom bombardment. Biochemist 23:6389–6392

Tran H, Tanaka A, Litvinovich SV, Medved LV, Haudenschild CC, Argraves WS (1995) The interaction of fibulin-1 with fibrinogen. A potential role in hemostasis and thrombosis. J Biol Chem 270:19458–19464

Tran R, Myers DR, Ciciliano J, Trybus Hardy EL, Sakurai Y, Ahn B, Qiu Y, Mannino RG, Fay ME, Lam WA (2013) Biomechanics of haemostasis and thrombosis in health and disease: from the macro- to molecular scale. J Cell Mol Med 17:579–596

Tsurupa G, Yakovlev S, Mckee P, Medved L (2010) Noncovalent interaction of alpha(2)-antiplasmin with fibrin(ogen): localization of alpha(2)-antiplasmin-binding sites. Biochemistry 49:7643–7651

Tsurupa G, Mahid A, Veklich Y, Weisel JW, Medved L (2011) Structure, stability, and interaction of fibrin alphaC-domain polymers. Biochemistry 50:8028–8037

Tsurupa G, Pechik I, Litvinov RI, Hantgan RR, Tjandra N, Weisel JW, Medved L (2012) On the mechanism of alphaC polymer formation in fibrin. Biochemistry 51:2526–2538

Uitte De Willige S, De Visser MC, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM (2005) Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. Blood 106:4176–4183

Uitte De Willige S, Standeven KF, Philippou H, Ariens RA (2009) The pleiotropic role of the fibrinogen gamma' chain in hemostasis. Blood 114:3994–4001

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:e88–e99

Undas A, Brozek J, Jankowski M, Siudak Z, Szczeklik A, Jakubowski H (2006) Plasma homocysteine affects fibrin clot permeability and resistance to lysis in human subjects. Arterioscler Thromb Vasc Biol 26:1397–1404

Valnickova Z, Enghild JJ (1998) Human procarboxypeptidase U, or thrombin-activable fibrinolysis inhibitor, is a substrate for transglutaminases. Evidence for transglutaminase-catalyzed cross-linking to fibrin. J Biol Chem 273:27220–27224

Varju I, Sotonyi P, Machovich R, Szabo L, Tenekedjiev K, Silva MM, Longstaff C, Kolev K (2011) Hindered dissolution of fibrin formed under mechanical stress. J Thromb Haemost 9:979–986

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 268:13577–13585

Veklich Y, Francis CW, White J, Weisel JW (1998) Structural studies of fibrinolysis by electron microscopy. Blood 92:4721–4729

Wei AH, Schoenwaelder SM, Andrews RK, Jackson SP (2009) New insights into the haemostatic function of platelets. Br J Haematol 147:415–430

Weigandt KM, White N, Chung D, Ellingson E, Wang Y, Fu X, Pozzo DC (2012) Fibrin clot structure and mechanics associated with specific oxidation of methionine residues in fibrinogen. Biophys J 103:2399–2407

Weisel JW (1986) The electron microscope band pattern of human fibrin: various stains, lateral order, and carbohydrate localization. J Ultrastruct Mol Struct Res 96:176–188

Weisel JW (2004) Cross-linked gamma chains in a fibrin fibril are situated transversely between its strands: no. J Thromb Haemost 2:394–399

Weisel JW (2005) Fibrinogen and fibrin. In: Parry DAD, Squire J (eds) Coiled-coils, collagen & elastomers. Elsevier, San Diego

Weisel JW (2007) Structure of fibrin: impact on clot stability. J Thromb Haemost 5(Suppl 1):116–124

Weisel JW, Dempfle C-EH (2013) Fibrinogen structure and function. In: Marder V, Aird WC, Bennett JS, Schulman S, White GC (eds) Hemostasis and thrombosis: basic principles and clinical practice, 6th edn. Lippincott Williams and Wilkins, Philadelphia

Weisel JW, Litvinov RI (2008) The biochemical and physical process of fibrinolysis and effects of clot structure and stability on the lysis rate. Cardiovasc Hematol Agents Med Chem 6:161–180

Weisel JW, Litvinov RI (2013) Mechanisms of fibrin polymerization and clinical implications. Blood 121:1712–1719

Weisel JW, Litvinov RI (2014) Mechanisms of fibrinolysis and basic principles of management. In: Saba HI, Roberts HR (eds) Hemostasis and thrombosis: practical guidelines in clinical management. Wiley-Blackwell, Chichester

Weisel JW, Medved L (2001) The structure and function of the alpha C domains of fibrinogen. Ann N Y Acad Sci 936:312–327

Weisel JW, Nagaswami C (1992) Computer modeling of fibrin polymerization kinetics correlated with electron microscope and turbidity observations: clot structure and assembly are kinetically controlled. Biophys J 63:111–128

Weisel JW, Warren SG, Cohen C (1978) Crystals of modified fibrinogen: size, shape and packing of molecules. J Mol Biol 126:159–183

Weisel JW, Phillips GJ, Cohen C (1983) The structure of fibrinogen and fibrin: II. Architecture of the fibrin clot. Ann N Y Acad Sci 408:367–379

Weisel JW, Stauffacher CV, Bullitt E, Cohen C (1985) A model for fibrinogen: domains and sequence. Science 230:1388–1391

Weisel JW, Veklich Y, Gorkun O (1993) The sequence of cleavage of fibrinopeptides from fibrinogen is important for protofibril formation and enhancement of lateral aggregation in fibrin clots. J Mol Biol 232:285–297

Weiss HL, Selvaraj P, Okita K, Matsumoto Y, Voie A, Hoelscher T, Szeri AJ (2013) Mechanical clot damage from cavitation during sonothrombolysis. J Acoust Soc Am 133:3159–3175

Wen Q, Janmey PA (2013) Effects of non-linearity on cell-ECM interactions. Exp Cell Res 319:2481–2489

Whittaker P, Przyklenk K (2009) Fibrin architecture in clots: a quantitative polarized light microscopy analysis. Blood Cells Mol Dis 42:51–56

Williams RC (1981) Morphology of bovine fibrinogen monomers and fibrin oligomers. J Mol Biol 150:399–408

Wolberg AS (2010) Plasma and cellular contributions to fibrin network formation, structure and stability. Haemophilia 16(Suppl 3):7–12

Wolberg AS (2012) Determinants of fibrin formation, structure, and function. Curr Opin Hematol 19:349–356

Wolfenstein TC, Mosesson MW (1981) Carboxy-terminal amino acid sequence of a human fibrinogen gamma-chain variant (gamma'). Biochemist 20:6146–6149

Wufsus AR, Rana K, Brown A, Dorgan JR, Liberatore MW, Neeves KB (2015) Elastic behavior and platelet retraction in low- and high-density fibrin gels. Biophys J 108:173–183

Yakovlev S, Medved L (2015) Interaction of fibrin with the very low density lipoprotein receptor: further characterization and localization of the fibrin-binding site. Biochemistry 54:4751–4761

Yakovlev S, Gao Y, Cao C, Chen L, Strickland DK, Zhang L, Medved L (2011) Interaction of fibrin with VE-cadherin and anti-inflammatory effect of fibrin-derived fragments. J Thromb Haemost 9:1847–1855

Yamazumi K, Doolittle RF (1992) The synthetic peptide Gly-Pro-Arg-Pro-amide limits the plasmic digestion of fibrinogen in the same fashion as calcium ion. Protein Sci 1:1719–1720

Yang Z, Mochalkin I, Doolittle LR (2000) A model for fibrin formation based on crystal structures of fibrinogen and fibrin fragments complexed with specific peptides. Proc Natl Acad Sci U S A 97:14156–14161

Yee VC, Pratt KP, Cote 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 5:125–138

Yermolenko IS, Lishko VK, Ugarova TP, Magonov SN (2011) High-resolution visualization of fibrinogen molecules and fibrin fibers with atomic force microscopy. Biomacromolecules 12:370–379

Yeromonahos C, Polack B, Caton F (2010) Nanostructure of the fibrin clot. Biophys J 99:2018–2027

Zalewski J, Bogaert J, Sadowski M, Woznicka O, Doulaptsis K, Ntoumpanaki M, Zabczyk M, Nessler J, Undas A (2015) Plasma fibrin clot phenotype independently affects intracoronary thrombus ultrastructure in patients with acute myocardial infarction. Thromb Haemost 113:1258–1269

Zhang JZ, Redman CM (1996) Fibrinogen assembly and secretion. Role of intrachain disulfide loops. J Biol Chem 95:30083–30088

Zhmurov A, Brown AE, Litvinov RI, Dima RI, Weisel JW, Barsegov V (2011) Mechanism of fibrin(ogen) forced unfolding. Structure 19:1615–1624

Zhmurov A, Kononova O, Litvinov RI, Dima RI, Barsegov V, Weisel JW (2012) Mechanical transition from alpha-helical coiled coils to beta-sheets in fibrin(ogen). J Am Chem Soc 134:20396–20402

Zhmurov A, Protopopova AD, Litvinov RI, Zhukov P, Mukhitov AR, Weisel JW, Barsegov V. (2016) Structural basis of interfacial flexibility in fibrin oligomers. Structure, 24:1907–1917

Zubairova LD, Nabiullina RM, Nagaswami C, Zuev YF, Mustafin IG, Litvinov RI, Weisel JW (2015) Circulating microparticles alter formation, structure, and properties of fibrin clots. Sci Rep 5:17611

Chapter 14
Fibrillar Collagens

Jordi Bella and David J.S. Hulmes

Contents

Abstract Fibrillar collagens (types I, II, III, V, XI, XXIV and XXVII) constitute a sub-group within the collagen family (of which there are 28 types in humans) whose functions are to provide three-dimensional frameworks for tissues and organs. These networks confer mechanical strength as well as signalling and organizing

J. Bella (✉)
Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK
e-mail: jordi.bella@manchester.ac.uk

D.J.S. Hulmes
Tissue Biology and Therapeutic Engineering Unit (UMR5305), CNRS/Université Claude Bernard Lyon 1, Lyon, France

functions through binding to cellular receptors and other components of the extracellular matrix (ECM). Here we describe the structure and assembly of fibrillar collagens, and their procollagen precursors, from the molecular to the tissue level. We show how the structure of the collagen triple-helix is influenced by the amino acid sequence, hydrogen bonding and post-translational modifications, such as prolyl 4-hydroxylation. The numerous steps in the biosynthesis of the fibrillar collagens are reviewed with particular attention to the role of prolyl 3-hydroxylation, collagen chaperones, trimerization of procollagen chains and proteolytic maturation. The multiple steps controlling fibril assembly are then discussed with a focus on the cellular control of this process *in vivo*. Our current understanding of the molecular packing in collagen fibrils, from different tissues, is then summarized on the basis of data from X-ray diffraction and electron microscopy. These results provide structural insights into how collagen fibrils interact with cell receptors, other fibrillar and non-fibrillar collagens and other ECM components, as well as enzymes involved in cross-linking and degradation.

14.1   Introduction

Collagens are characteristic proteins of the metazoan family. Their main function is to provide mechanical stability to animals, from sponges to higher vertebrates, and therefore they form the most abundant group of proteins in connective tissues (Myllyharju and Kivirikko 2004). Several collagen types have been described at the genetic level; there are more than 40 collagen genes in the human or mouse genomes, which combine to form up to 28 collagen types. These are designated with roman numerals I to XXVIII, following the order of their discovery. Additionally, a number of proteins structurally related to collagens are classified as collagen-like proteins, taking a typical vertebrate *collagenome* (human, mouse) to more than 80 collagen and collagen-like genes (Kadler et al. 2007; Gordon and Hahn 2010; Ricard-Blum 2011; Mienaltowski and Birk 2014). Most collagens participate in some form of higher-order supramolecular self-assembly that directly relates to their biological function. Interestingly, collagens appear to be largely resistant to and have a protective effect against formation of self-assembled toxic aggregates such as amyloid fibrils (Parmar et al. 2012; Dubey and Kar 2014).

14.2   Fibrillar and Non-fibrillar Collagens

Vertebrate collagens can be classified into two broad categories based on their supramolecular organization: fibrillar and non-fibrillar. The main fibril-forming collagens are types I, II and III (major fibrillar collagens) and also V and XI (minor fibrillar collagens). They form elongated, cable-like striated fibril structures with a distinct 64–67-nm band periodicity and have a clear structural role of mechanical support and dimensional stability. Two additional fibrillar collagen types, XXIV and XXVII, have been described, but these are much less abundant and their function is still being investigated (Exposito et al. 2010). Non-fibrillar collagens may form other types of supramolecular structures: networks of different topologies (collagens IV, VIII and X), beaded filaments (VI, XXVI and XXVIII) or anchoring fibrils (VII). Some non-fibrillar collagens are associated with the surface of collagen fibrils (IX, XII, XIV, XIV and XIX-XXII) and some are transmembrane proteins (XIII, XVII, XXIII and XXV) (Myllyharju and Kivirikko 2004; Kadler et al. 2007; Gordon and Hahn 2010; Ricard-Blum 2011; Mienaltowski and Birk 2014). Fibrillar collagens are differently distributed across connective tissues. Thus, type I collagen is the main protein in bone, skin or tendon, type II collagen is specific to cartilage, and type III collagen, together with type I, are the main structural components of blood vessels. Collagen proteins in invertebrates have similar structural roles. For instance, the cuticle of nematodes is an exoskeleton made of cuticle collagen proteins which are encoded by more than 170 genes (Page and Johnstone 2007).

14.3   Fibrillar Collagen Genes and Polypeptide Chains

All collagens are trimeric proteins where three polypeptide chains must come together to form a functional collagen molecule. Each individual type of chain is encoded by a different gene and is named with the Greek letter α followed by the chain number and its collagen type. Thus, the human *Col1a1* and *Col1a2* genes code for the human α1(I) and α2(I) chains, respectively, the mouse *Col11a1* gene codes for the mouse α1(XI) chain, and so forth. A single exception seems to be the α3(XI) chain, which is encoded by the *Col2a1* gene (see below). The entire complement of fibrillar collagen chains in humans or mouse genomes is encoded by eleven genes (Table 14.1). Phylogenetic analysis of these collagen genes indicates the existence of three clearly defined clades or subclasses named A, B and C (Boot-Handford and Tuckwell 2003; Huxley-Jones et al. 2007; Exposito et al. 2010).

The three chains building a collagen molecule can be identical (homotrimer collagens) or different (heterotrimer collagens). The major fibrillar collagens I, II and III have a rather defined composition: types II and III are homotrimers of three identical polypeptide chains, $[α1(II)]_3$ and $[α1(III)]_3$, respectively, whereas most vertebrates build type I collagen as heterotrimers of two different chains α1(I) and α2(I), with an $[α1(I)]_2α2(I)$ stoichiometry (Table 14.2).

Table 14.1 Fibrillar collagen genes in human, mouse and zebrafish genomes. NCBI and UniProt accession numbers and numbers of amino acids (Naa). Many genes are duplicated in fish genomes, and the two copies are usually distinguished with letters a and b

Gene	Chain	Clade	Human	Naa	Mouse	Naa	Zebrafish	Naa
Col1a1, Col1a1a	α1(I), α1(I)a	A	NP_000079 P02452	1464	NP_031768 P11087	1453	NP_954684 Q6P4U1	1447
Col1a2	α2(I)	A	NP_000080 P08123	1366	NP_031769 Q01149	1372	NP_892013 Q6IQX2	1352
Col1a3, Col1a1b	α3(I), α1(I)b	A					NP_958886 Q6PEI9	1449
Col2a1, Col2a1a	α1(II), α1(II)a	A	NP_001835 P02458	1487	NP_112440 P28481	1487	NP_571367 Q2LDA1	1491
Col2a1b	α1(II)b	A					NP_001268407 L0S5L0	1493
Col3a1	α1(III)	A	NP_000081 P02461	1466	NP_034060 P08121	1464	Absent	
Col5a1	α1(V)	B	NP_000084 P20908	1838	NP_056549 O88207	1838	XP_009293460 F6NPA4	2069
Col5a2, Col5a2a	α2(V), α2(V)a	A	NP_000384 P05997	1499	NP_031763 Q3U962	1497	NP_001139254 F1QT86	1491
Col5a2b	α2(V)b	A					XP_001920005 E7FE47	1499
Col5a3, Col5a3a	α3(V), α3(V)a	B	NP_056534 P25940	1745	NP_058615 Q9JLI2	1739	NP_001177685 D6MUD6	1887
Col5a3b	α3(V)b	B					NP_001177687 D6MUD7	1723
Col11a1, Col11a1a	α1(XI), α1(XI)a	B	NP_001845 P12107	1806	NP_031755 Q61245	1804	NP_001077313 A2RUX5	1736

Col11a1b	α1(XI)b	B					NP_001171883 D6MUD3	1815
Col11a2	α2(XI)	B	NP_542411 P13942	1736	NP_034056 Q64739	1736	NP_001315194 D6MUD4	1803
Col24a1	α1(XXIV)	C	NP_690850 Q17RW2	1714	NP_082046 Q30D77	1733	XP_002666614 F8W5B3	201[a]
Col27a1, *Col27a1a*	α1(XXVII), α1(XXVII)a	C	NP_116277 Q8IZC6	1860	NP_079961 Q5QNQ9	1845	NP_001156766 C7DZK3	1783
Col27a1b	α1(XXVII)b	C					NP_001074044 A0MSJ1	1658

[a]Fragment sequence

Table 14.2 Chain compositions of human fibrillar collagens (Eyre et al. 2006; Ricard-Blum 2011)

Collagen types	Chains	Compositions
Collagen I	α1(I), α2(I)	[α1(I)]$_2$ α2(I)
Collagen II	α1(II)	[α1(II)]$_3$
Collagen III	α1(III)	[α1(III)]$_3$
Collagen V	α1(V), α2(V), α3(V)	[α1(V)]$_3$
		[α1(V)]$_2$α2(V)
		α1(V)α2(V)α3(V)
Collagen XI	α1(XI), α2(XI), α3(XI)	α1(XI)α2(XI)α3(XI)
Heterotypic collagen V/XI	α1(V), α2(XI), α3(XI)	α1(XI)α1(V)α3(XI)
		α1(V)α1(XI)α2(V)
		[α1(XI)]$_2$α2(V)

Chains from types V and XI can form heterotypic trimeric collagen molecules

Interestingly, a third chain for type I collagen, α3(I), has been described in the genomes of several teleostean fish, although it is not currently known how this chain associates with the other two, and what new combinations of type I collagen chains may occur in bony fish. The fish-specific gene for α3(I) has been recently linked to thermal remodelling of cardiac tissue in rainbow trout, suggesting possible implications for the different pathological response to fibrosis in fish and mammals (Keen et al. 2016). The *Col1a3* gene for the α3(I) chain is considered to result from a duplication of the *Col1a1* gene in fish, and the two duplicates have been renamed *Col1a1a* and *Col1a1b* in the zebrafish genome (Table 14.1). Zebrafish and other teleostean fish genomes show duplication of several other fibrillar collagen genes, while the gene for type III collagen *Col3a1* is absent (Duran et al. 2015a). Duplicate fibrillar collagen genes may offer additional possibilities for regulation and in zebrafish they have been shown to be expressed differentially during fin skeleton regeneration (Duran et al. 2015a). Some of the duplicate genes may replace type III collagen function in tissues common to all vertebrates, such as vasculature or skeleton.

Minor fibrillar collagen types V and XI can have different chain compositions. Thus, type V collagen in humans has at least three different isoforms: [α1(V)]$_3$, [α1(V)]$_2$α2(V) or α1(V)α2(V)α3(V), with different tissue distributions and biological roles (Mak et al. 2016). Type XI collagen appears more complicated, as the α3(XI) chain is in fact encoded by the *Col2a1* gene and the composition of heterotrimeric collagen XI has been described either as α1(XI)α2(XI)α3(XI) or α1(XI) α2(XI)α1(II). It seems that the first description should be preferred as α3(XI) is processed differently from the α1(II) chain forming homotrimer type II collagen (Thom and Morris 1991). Furthermore, types V and XI are able to form tissue-specific heterotypic trimeric molecules with chains from the two different types (Table 14.2). Thus, these two collagens should not be considered distinct types but rather a single collagen type V/XI with multiple isoforms (Fichard et al. 1995; Smith and Birk 2012; Mak et al. 2016). The number of trimeric proteins for collagen types V and XI may increase through alternative splicing of several collagen

genes. These alternatively spliced variants are produced in different tissues or at different stages of development (Hoffman et al. 2010; Fang et al. 2012; McAlinden 2014).

Comparatively little is known about the biological functions of the more recently discovered collagen types XXIV and XXVII (Boot-Handford et al. 2003; Koch et al. 2003). They show significant structural differences with the other fibrillar collagens and form their own subclass (clade C). They are thought to be homotrimers and their fibrils seem to be non-striated and thinner than those of major fibrillar types (Plumb et al. 2007).

The evolution of the fibrillar collagen family from sponges to mammals has been extensively studied and an evolutionary model spanning the entire metazoan lineage has been proposed on the basis of phylogenetic data (Exposito et al. 2010). Collagens are ancient molecules: collagen sequences are already found in choanoflagellate genomes and it seems that the ancestral fibrillar collagen gene appeared before the metazoan expansion. Early duplication events led to the formation of the three A, B and C clades, before the emergence of chordates and probably predating the split between radiata and bilateria (Exposito et al. 2008).

14.4 Domain Structures of Fibrillar Collagen Chains

Collagens are multidomain proteins. Each collagen chain contains one or more instances of a characteristic triple-helical domain (COL), flanked by or interspersed with other non-collagenous domains (NC). The usual convention is to name the different domains starting from the C-terminus as NC1, COL1, NC2, COL2, and so forth. Different collagen types have different numbers and arrangements of COL and NC domains, and the primary sequences of their polypeptide chains determine which collagen type they belong to (Gordon and Hahn 2010; Ricard-Blum 2011). Fibrillar collagens, in particular, have in common a long triple helical COL1 domain (~1014–1020 amino acids, about 300 nm long) and a specific C-terminal NC1 domain (or C-propeptide) that is also trimeric (Fig. 14.1).

The N-terminal regions (N-propeptides) can have different domain architectures that correspond to the different clades of the fibrillar collagen genes. N-propeptides usually contain one of more short segments of triple-helical domain sequences (the minor triple-helical region), with the longest uninterrupted segment being 40–60 amino acids long. This is often termed the minor helix or alternatively the COL2 domain (Fig. 14.1). The fate of this minor helix during collagen processing (see below) depends on the collagen type. In clade A collagen types I, II and III, the N-propeptide is removed completely by endopeptidases that cleave the three chains between the minor (COL2) and major (COL1) triple helices. Together with the endopeptide cleavage at the other end, processed collagens I, II and III (sometimes called tropocollagens) are essentially long COL1 domains with short non-triple helical sequences at both ends named telopeptides. These short telopeptides provide sites for crosslinking during fibril formation.
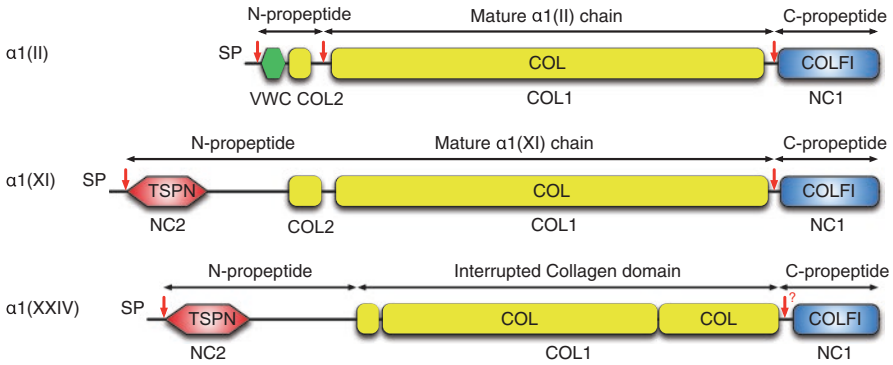
Fig. 14.1 Domain structures of representative examples of fibrillar collagen α chains: α1(II) as a clade A major fibrillar collagen, α1(XI) as a clade B minor fibrillar collagen, and α1(XXIV) as a clade C fibrillar collagen. Domain names: *COL* collagen triple helical domain, *COLFI* C-terminal domain from fibrillar collagens, *TSPN* thrombospondin N-terminal domain, also known as laminin globular (G) domain (*LamG*); *VWC* von Willebrand factor type C domain, *SP* signal peptide. Collagen chains are often described as made of non-collagen (*NC*) and collagen (*COL*) domains numbered starting from the C-terminus. Clades A and B have a minor collagen domain (*COL2*) with interruptions and imperfections in its sequence. *Red arrows* indicate sites of proteolytic cleavage, in the case of α1(XXIV) this is hypothesized even though the actual site has not been identified yet

Most clade A collagens also have a short von Willebrand factor type C module (VWC) in their N-propeptide region (Fig. 14.1). This module is removed in the mature chains as part of collagen processing. N-propeptides of collagens from clades B and C contain a domain with homology to the terminal domain of thrombospondin (TSPN; Fig. 14.1), which is also observed in several non-fibrillar collagen sequences. This domain type has been recently renamed the laminin globular (LamG) domain in protein databases, and different literature sources may use either of the two domain names when describing the architecture of collagen N-propeptides. Processing of clade B collagens usually involves cleaving the C-propeptide while the N-propeptide is partially retained. Thus, the mature forms of collagen types V and XI usually have two or three chains with retained N-propeptides containing a TSPN/LamG domain and a minor collagen helix, separated from the major collagen helix by a short flexible region. This structure is important for the function of collagen V/XI types as regulators of fibril diameter: the retained N-propeptides are thought to project outwards from the surface of the growing fibril, limiting further growth via steric hindrance (see below).

Clade C collagens XXIV and XXVII show some significant structural differences with the other two clades. They have a slightly shorter COL1 domain (less than 1000 amino acids) that contains two interruptions. Their N-propeptide regions do not contain a minor triple helix and they are supposed to remain attached to the COL1 domain after processing. It is expected that the C-propeptide regions of collagen types XXIV and XXVII are removed during processing to the mature forms

of these collagens, although the cleavage sites have not yet been identified. The combination of interrupted COL1 domains and retained N-propeptide regions is probably the reason for the very thin fibrils observed for collagen type XXVII (Plumb et al. 2007). Further structural details about these collagens will surely emerge as the research into their biological function progresses.

14.5 Molecular Structure and Stability of the Collagen Triple Helix

The collagen triple helix (COL domain) is the common structural motif of all collagen proteins (Fig. 14.2: see Chap. 1). At the amino acid sequence level this motif is easily recognized by a repetitive pattern (Gly-Xaa-Yaa)$_n$, where imino acid residues often occupy the Xaa and Yaa positions. The average structure of the collagen
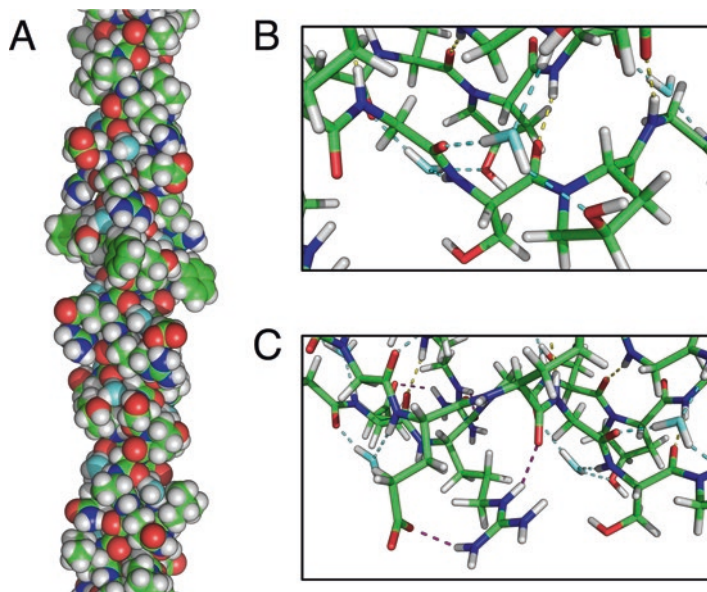
Fig. 14.2 Model of the collagen triple helix of human type II collagen around the integrin binding sequence GFOGER. Water molecules mediating inter-strand hydrogen bonding have been modelled in place. Side-chains and hydrogen atoms have been modelled in idealized positions and orientation for hydrogen bonding connectivity. (**a**) Space-filling representation with carbon atoms in *green*, oxygen atoms from collagen chains in *red*, oxygen atoms from water molecules in *cyan*, nitrogen atoms in *blue* and hydrogen atoms in *white*. (**b**) Stick representation of main chain hydrogen: N–H···O=C hydrogen bonds (*yellow dashed*); water-mediated hydrogen bonds (*cyan dashed*); 4Hyp side-chains provide additional hydrogen bonding to the water molecules. (**c**) Stick representation of hydrogen bonds involving an arginine side-chain (*purple dashed*), to a main chain C=O group and to a glutamate side chain from a different strand

Table 14.3 Residue preferences (%) in sequences of COL domains from collagen types I-XXVIII in human, mouse and zebrafish genomes (O = one-letter code for 4-hydroxyproline, 4Hyp; standard one-letter codes for the rest of amino acids)

	Human collagens		Mouse collagens		Zebrafish collagens	
	Fibrillar	Non-fibrillar	Fibrillar	Non-fibrillar	Fibrillar	Non-fibrillar
Residues	Xaa/Yaa	Xaa/Yaa	Xaa/Yaa	Xaa/Yaa	Xaa/Yaa	Xaa/Yaa
P/O[a]	32/34	25/42	31/34	25/41	30/30	24/36
ED	18/7	20/5	19/7	20/6	18/7	20/6
KR	6/21	6/22	6/21	7/23	7/21	7/22
VILMF	18/11	26/9	18/11	26/9	18/12	25/9
A	10/9	6/5	9/9	5/5	10/10	5/5
TS	6/8	8/6	8/8	8/7	7/10	7/9
QN	6/9	6/7	6/8	6/8	6/9	7/9
G	1/1	1/2	2/1	1/2	2/1	1/3
HYCW[b]	2/1	3/2	2/1	3/2	2/1	3/2

[a]Statistics refer to Pro in the Xaa position and 4Hyp in the Yaa position assuming an ideal case of 100 % hydroxylation

[b]Cys and Trp residues are almost completely absent from the COL domains of fibrillar collagens

triple helix was initially determined from fibre X-ray diffraction of tendons, which are primarily built from type I collagen. Introduction at a later stage of chemically synthesized collagen-like peptides allowed for a more precise characterisation of the molecular details, plus a better understanding of the mechanisms of thermal stability of the triple helix. A summary of our current knowledge follows below. A comprehensive review on the triple-helical collagen structure has been recently published (Bella 2016) and the reader is referred to it for a more detailed description.

14.5.1 Overall Chain Conformation and Hydrogen Bonding

The three individual collagen chains adopt a helical conformation known as poly-proline II. They wrap around each other forming a one-residue staggered helical structure, which is held together by a ladder of inter-strand hydrogen bonds in a direction roughly perpendicular to the helical axis. This particular structure places every third residue on each chain close to the common helical axis, and is only ste-reochemically possible if the smallest amino acid (glycine) occupies this position, therefore resulting in the (Gly-Xaa-Yaa)$_n$ repetitive pattern of the COL domains. There are two main topologies for inter-strand hydrogen bonding (Fig. 14.2b). The N–H groups from the Gly residues at every third position form hydrogen bonds with the C=O groups of the amino or imino acids in the Xaa position on the next chain. A second set of water-mediated hydrogen bonds occurs between the N–H groups of the amino acids in the Xaa position and the C=O groups from glycine residues in the previous chain. The first set of hydrogen bonds occurs once per every Gly-Xaa-Yaa triplet; the second set is only possible when Xaa is not an imino acid, about 70 % of the times in major fibrillar collagens (Table 14.3). Overall, about 1.7 N–H groups

per every three residues in a COL domain are involved in inter-strand hydrogen bonding.

14.5.2 4-Hydroxyproline and Collagen Stability

Compared to most other proteins, the amino acid composition of the vertebrate collagen proteins is highly unusual. Collagens have an unusually high proportion of imino acids, close to 30 %, in the sequences of their COL domains. Proline (Pro, P) is the most common residue in the Xaa position, while its derivative 4(*R*)-hydroxyproline (4Hyp, O) is the most common residue in the Yaa position (Table 14.3). The high content of 4Hyp is a defining characteristic of collagens. This imino acid is formed through post-translational modification of proline (Pro) residues in the Yaa position by one of three (in humans) collagen prolyl 4-hydroxylases (C-P4Hs), tetrameric enzymes of two α chains and two β chains, the β chain being protein disulphide isomerase (Myllyharju 2008; Gorres and Raines 2010). Prolyl hydroxylation is essential for animal life, and removal of C-P4H activity has proven lethal for vertebrate and invertebrate animal models. The C-P4H enzyme uses atmospheric O_2 and is very specific: only Pro residues in the Yaa position of individual polypeptide chains with repeating Gly-Xaa-Yaa sequence become hydroxylated to 4Hyp; those in the Xaa position remain unhydroxylated. Furthermore, C-P4H only acts on Pro prior to formation of collagen triple helices. 4(*R*)-hydroxylation of Pro residues in the Yaa position increases the thermal stability of the collagen triple helix to levels compatible with the body temperature of the animal. While determination of a complete C-P4H structure has proven elusive so far, a recent crystal structure of part of the α subunit of C-P4H-I has been reported. It reveals a dimer held together by a coiled-coil in the N-terminal domain and a peptide-substrate-binding domain characterised by a tyrosine-rich groove (Anantharajan et al. 2013; Fig. 14.3a).

Both Pro and 4Hyp residues stabilize the collagen triple helix through stereochemical restrictions imposed by their imino acid rings. This is a largely entropic effect: imino acids decrease the entropic cost of formation of the triple helix by favouring polyproline II-like conformations in the unfolded state (Vitagliano et al. 2001; Bella 2016). Importantly, changing Pro residues in the Yaa position to 4Hyp has an additional stabilizing effect due to the preferred conformation of the 4Hyp imino acid ring induced by stereoelectronic effects, and the impact that the ring torsion angles have on the main chain conformation (the *propensity-based hypothesis* for collagen stabilization, see (Vitagliano et al. 2001; Shoulders and Raines 2009; Shoulders et al. 2010; Bella 2016)). In addition, 4Hyp residues in the Yaa position provide additional stabilization, this time of enthalpic nature, through interaction with water molecules surrounding the triple helix (Nishi et al. 2005). In particular, 4Hyp residues provide additional hydrogen bonding to water molecules involved in
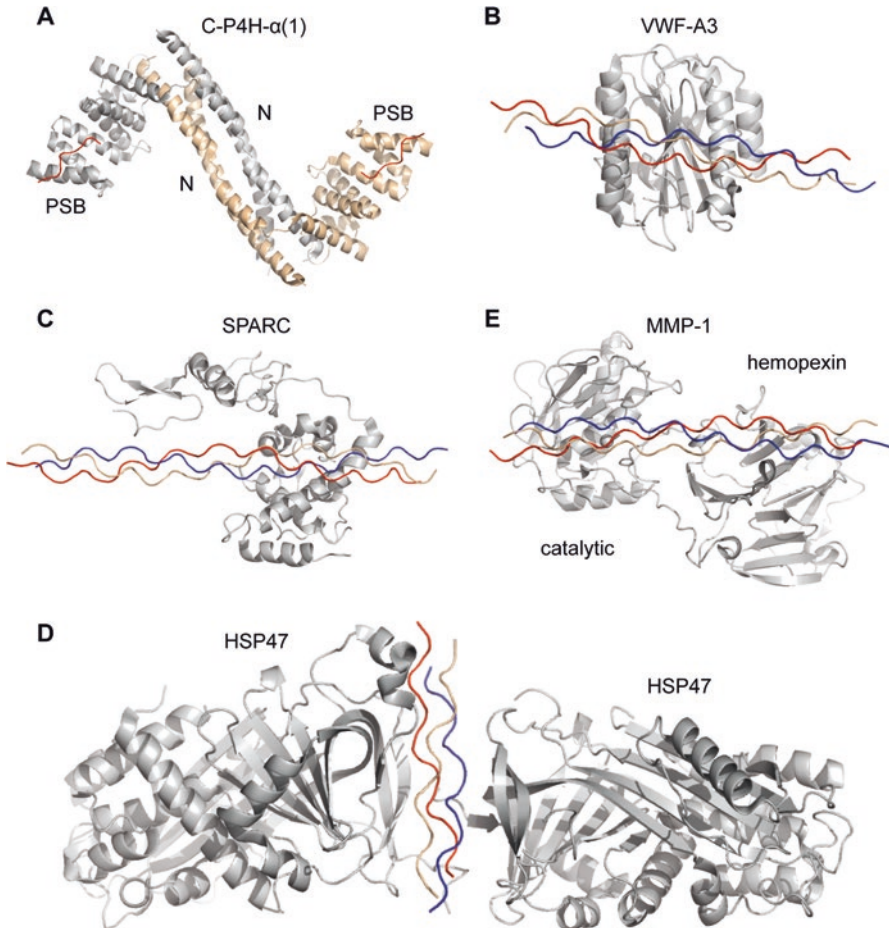
Fig. 14.3 Crystal structures of collagen in complex with different interaction partners. (**a**) N-terminal region of the α subunit of collagen prolyl 4-hydroxylase-1 showing dimerization via the N domains (N) and binding of single-chain peptides (*red*) to the substrate binding domains (*PSB*). From (Anantharajan et al. 2013). (**b**) Complex of the von Willebrand factor A3 domain with a collagen triple-helical peptide (*THP*). (From Brondijk et al. 2012). (**c**) Binding of a *THP* to SPARC/osteonectin/BM-40. (From Hohenester et al. 2008). (**d**) Binding of two molecules of HSP47 to a single *THP*. (From Widmer et al. 2012). (**e**) Binding of a *THP* substrate to the catalytic and hemopexin domains of matrix metalloproteinase-1. (From Manka et al. 2012). For the triple-helical peptides, the trailing, middle and leading strands (original definition) are coloured *red*, *beige* and *blue*, respectively (Figures generated using PyMol)

the water-bridged inter-strand hydrogen bonding topology (Fig. 14.2b). The enthalpic stabilisation by 4Hyp residues is partially counteracted by the entropic penalty of organizing water molecules around the collagen triple helix, but the overall balance seems to be favourable for triple helix formation (Nishi et al. 2005; Bella 2016).

14.5.3 Additional Mechanisms of Collagen Stabilization

Other amino acid types have clear preferences for either the Xaa or Yaa positions of COL domains, or they may be relatively rare on COL sequences altogether (Table 14.3). Such amino acid positional preferences are often related to additional mechanisms of stabilisation of the collagen triple helix, as shown by the differences in thermal stability of collagen model peptides incorporating specific amino acids at either the Xaa or Yaa position (Persikov et al. 2005). Plausible explanations at the molecular level have been proposed for some cases. For instance, charged side chains of opposite sign can form inter-strand salt-bridges that stabilize the triple-helical over the single-chain form (Fallas et al. 2012). These interactions have a more favourable geometry when the negatively charged residue (Asp, D; Glu, E) is in the Xaa position and the positively charged one (Lys, K; Arg, R) is in the Yaa position (Fig. 14.2c). This preference is reflected in the amino acid composition of COL domains in vertebrate collagens (Table 14.3). Arginine (Arg, R) and glutamine (Gln, Q) side chains on one strand can form hydrogen bonds to the main chain on the next strand, also stabilizing the triple helix (Kramer et al. 2001; Fig. 14.2c). Arginine is relatively frequent in the Yaa position of vertebrate COL domains, whereas glutamine residues often occupy the Yaa position in COL domains of collagen-like proteins from prokaryotes, where stabilization via prolyl hydroxylation is not possible (Ghosh et al. 2012; Bella 2016). Some invertebrate collagens and some prokaryotic collagen-like proteins are stabilized by a completely different mechanism in which threonine (Thr, T) residues in the Yaa position are glycosylated (Bann et al. 2000, 2003; Daubenspeck et al. 2004). This specific mechanism is consistent with a high proportion of Thr residues in the Yaa position of their COL domain sequences (Bella 2016), which is not observed in vertebrate collagens. The combination of the different mechanisms of stabilization ensures that the denaturation temperature of animal collagens is close to the normal body temperature of the particular animal (Leikina et al. 2002; Persikov and Brodsky 2002).

14.5.4 Helical Twist of the Collagen Triple Helix

High resolution crystal structures of collagen model peptides suggest that the degree of triple helical twist changes with the proportion of imino acids Pro, Hyp in the sequence (Bella 2010, 2016). Regions with high imino acid content are slightly more twisted, forming a triple helix with roughly 3.5 residues per turn. The triple helix relaxes in regions devoid of imino acids and approaches 3.3 residues per turn. When the entire sequence of a fibrillar COL domain is considered, an average value of 3.4 residues per turn is obtained (Bella 2016). The actual conformation of the triple helix is likely to change along the sequence due to these intrinsic preferences, but also to the local arrangement and lateral interactions of the individual collagen triple helices when forming part of larger assemblies such as fibrils (Orgel et al. 2014).

14.6   Chain Register in the Collagen Triple Helix

In any trimeric protein with threefold symmetry, its three subunits or polypeptide chains should be indistinguishable. In collagen, however, the one-residue stagger between strands makes them topologically non-equivalent, even in homotrimeric collagen triple helices (Bella 2016). This issue becomes important when describing the interaction of collagen with cell-surface receptors or collagen-binding proteins (Emsley et al. 2000; Hohenester et al. 2008; Brondijk et al. 2012; Widmer et al. 2012), or when heterotrimeric collagens are considered; the local environment and three-dimensional structure of a particular sequence motif depends on the mutual arrangement of the three chains forming the triple helix. Thus, it is important to elucidate the specific chain register of heterotrimer fibrillar collagens such as type I collagen or type V/XI collagen. The situation is even more complicated in non-fibrillar heterotrimeric collagens such as type IV collagen, where interruptions of different lengths on each chain occur at a given site. These *incommensurate interruptions* may well result in triple helical structures with different chain register at either side of the interruption (Bella 2014), and in such cases the chain register would need to be determined locally for each triple-helical segment of interest.

A useful notation to distinguish the three strands as *trailing*, *intermediate* and *leading* was introduced to describe the different interactions between each of the three collagen strands and an integrin receptor domain (Emsley et al. 2000). In its original inception the trailing and leading chains were defined as those with over-hangs at the N-terminus and C-terminus, respectively (the middle chain has no over-hangs) (Bella 2016). Unfortunately, this notation is not free from ambiguity and in fact the opposite convention has been used in some reports.

The experimental evidence on the chain register of heterotrimer fibril collagens is surprisingly sparse. Type I collagen has three possibilities, $\alpha 1 \alpha 1 \alpha 2$, $\alpha 1 \alpha 2 \alpha 1$ and $\alpha 2 \alpha 1 \alpha 1$, where the first chain in the list is the trailing strand, the second the middle strand, and the third the leading strand (using the convention above). Heterotrimeric types where the three chains are all different (types V/XI) have in principle six possibilities. Whilst no information is available for the chain register in collagen types V/XI, the $\alpha 1 \alpha 2 \alpha 1$ and $\alpha 2 \alpha 1 \alpha 1$ registers have been proposed for type I collagen on the basis of molecular modelling and densitometric scans of electron micrographs (Hofmann et al. 1978; Bender et al. 1982). No direct experimental determination of the chain register for type I collagen is yet available, but stereochemical considerations from the crystal structures of homotrimer collagen peptides in complex with either a von Willebrand factor A3 domain (Fig. 14.3b) or an integrin I-domain support the $\alpha 1 \alpha 1 \alpha 2$ register (Brondijk et al. 2012). Recent progress on the synthesis of heterotrimeric collagen model peptides with defined register (Acevedo-Jake et al. 2016) should provide the biochemical tools for confirming the $\alpha 1 \alpha 1 \alpha 2$ register for type I collagen and determining the chain register for other heterotrimer collagens.

14.7   Fibrillar Collagen Biosynthesis

All fibrillar collagens are initially produced as soluble biosynthetic precursors, pro-collagens. Inside the cell, numerous steps control the assembly of each procollagen molecule from its three constituent polypeptide chains (Ishikawa and Bachinger 2013). These include translation and transport of chains into the rough endoplasmic reticulum, hydroxylation of prolines and lysines, O- and N-linked glycosylation, chain association and folding, and transport to the Golgi and secretion into the extracellular matrix. Several enzymes and other proteins (mostly collagen-specific) are involved, including prolyl 3- and 4-hydroxylases, lysyl hydroxylases, galacto-syl- and glucosyl-galactosyl-transferases, protein disulphide isomerase, peptidyl prolyl cis-trans isomerases, molecular chaperones (particularly the collagen-specific HSP47) and trafficking proteins such as TANGO1 (D'Arcangelo et al. 2013). Once the procollagen molecules are formed, they are processed to collagen by proteolytic release of the N- and C-propeptides. These steps take place either during or after secretion and involve proteinases such as ADAMTSs (Bekhouche and Colige 2015) and bone morphogenetic protein-1/tolloid-like proteinases (BTPs) (Vadon-Le Goff et al. 2015). Mature collagen molecules then associate by D-staggered interactions to form fibrils, as regulated by cell-surface and extracellular proteins such as other collagens, integrins, fibronectin, SPARC (Fig. 14.3c) and small leucine-rich proteo-glycans (Hohenester et al. 2008; Kadler et al. 2008; Bruckner 2010; Kalamajski and Oldberg 2010; Sun et al. 2011; Chen and Birk 2013; Rosset and Bradshaw 2016). Due to space limitations, here we will present mainly recent advances in these areas. For earlier reviews, see (Myllyharju 2005; Hulmes 2008a, b; Ricard-Blum 2011; Yamauchi and Sricholpech 2012; Ishikawa and Bachinger 2013).

14.7.1   Role of 3-Hydroxyproline in Fibrillar Collagens

One area in which there has been considerable progress in recent years is the role of prolyl 3-hydroxylation in collagen biosynthesis and assembly. This second post-translational proline modification, 3(S)-hydroxyproline (3Hyp), was discovered more than 50 years ago in fibrillar collagens but its function is just being elucidated (Hudson and Eyre 2013). It is a rare modification, with only up to six occurrences per fibrillar collagen chain depending on collagen type, tissue, species and age (Weis et al. 2010; Hudson and Eyre 2013; Hudson et al. 2014; Taga et al. 2016). This contrasts with the far more abundant and regularly distributed 4Hyp (Table 14.2). However, 3Hyp sites appear to be crucial. They occur through post-translational modification of Pro residues in the Xaa position of a few specific (and conserved) Gly-Xaa-4Hyp triplets by collagen prolyl-3-hydroxylases (C-P3Hs) which are different from the C-P4Hs mentioned earlier. There are three C-P3Hs in humans. While the C-P4Hs are tetramers of two α chains and two β chains (the β chain being protein disulphide isomerase), C-P3Hs occur in trimeric complexes in
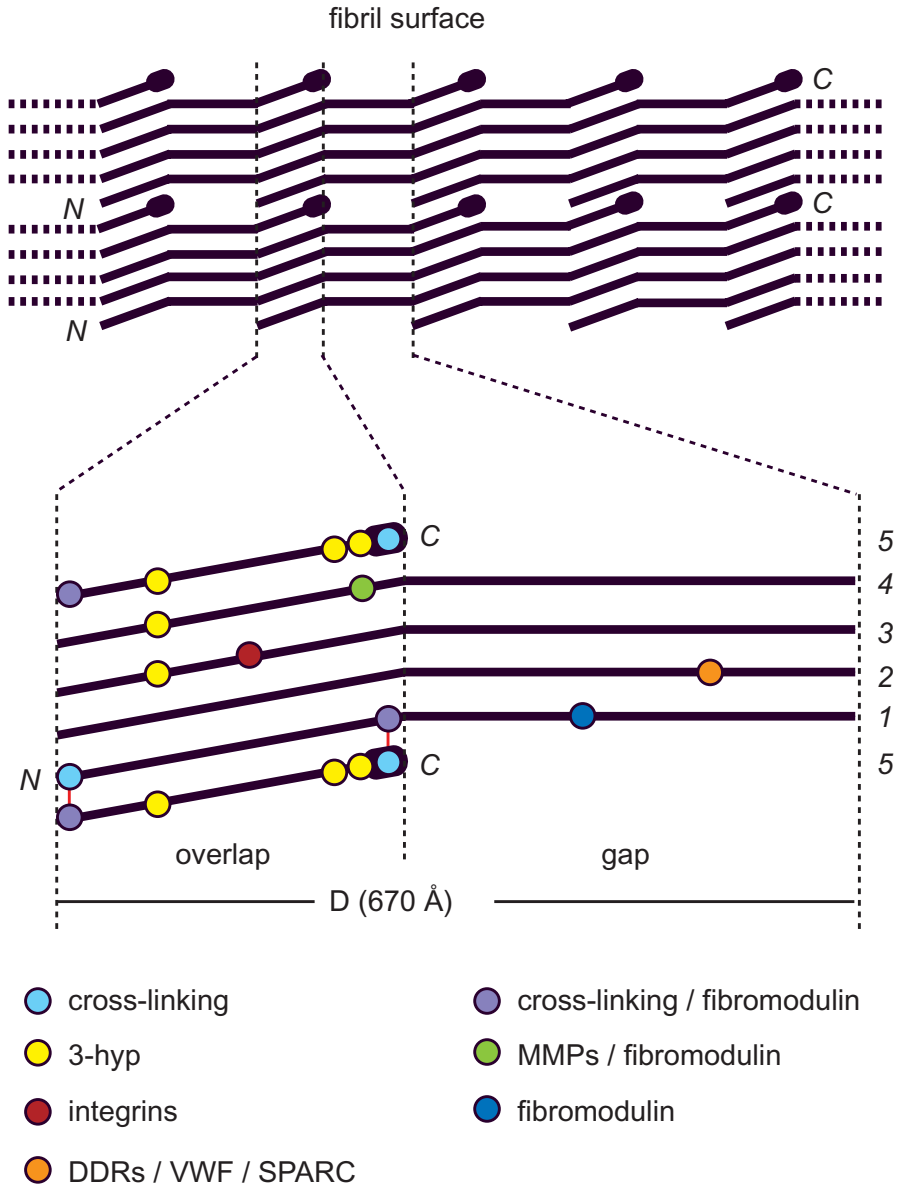
Fig. 14.4 Simplified diagram of the molecular packing at the collagen fibril surface with a zoom showing sites for intra-fibrillar cross-linking (indicated by *vertical red line*) and interactions with selected binding partners. According to structural data (Orgel et al. 2006; Perumal et al. 2008), the bulky C-telopeptide (*black oval*), is at the surface with the N-terminus buried in the fibril interior. Relative to the fibril axis, molecules are tilted in the overlap region but parallel in the gap region. Cross-linking and interaction sites are indicated, with data taken from (Orgel et al. 2006, 2011a, b; Sweeney et al. 2008; Hudson and Eyre 2013; Brondijk et al. 2012; and Kalamajski et al. 2016)

association with two other proteins of the rough endoplasmic reticulum, CRTAP (cartilage associated-protein) and cyclophilin B (CypB or PPIB; a peptidyl prolyl cis/trans isomerase). C-P3Hs are encoded by the *LEPRE* genes, and mutations in all three components of the C-P3H complex have been shown to give rise to recessive forms of the brittle bone disease osteogenesis imperfecta (Hudson and Eyre 2013). The reasons for this are unclear, but it is intriguing that some 3Hyp sites are found either close to cross-linking sites (see below) or at multiples of the D repeat (Fig. 14.4). Crystal structures of synthetic triple-helical peptides show that 3Hyp residues in the Xaa position are relatively exposed with their hydroxyl groups pointing away from the triple helix, unlike 4Hyp in the Yaa position (Fig. 14.2), and can form hydrogen bonds with neighbouring molecules (Schumacher et al. 2006). Whether the lack of such interactions in patients with deficient proline 3-hydroxylation is the cause of the observed phenotype is difficult to discern. Nevertheless, the low numbers of 3Hyp per chain suggest that its function is not primarily related to collagen stability but rather to provide sites of molecular recognition that are important for the self-assembly of fibrillar collagens and possibly for the interaction with other matrix proteins. Alterations in prolyl 3-hydroxylation could trigger further changes in collagen modifications (lysine hydroxylation, O-glycosylation, cross-linking) or induction of an unfolded protein response, all of which could lead to a change in phenotype (Vranka et al. 2010; Cabral et al. 2014).

14.7.2   Collagen Chaperones

Cyclophilin B, a component of the C-P3H complex, is but one of many peptidyl prolyl cis/trans isomerases involved in collagen biosynthesis (Ishikawa et al. 2015). Often these proteins also act as molecular chaperones, such as FKBP65 and FKBP22, both of which bind to collagens, the latter showing specificity (among the fibrillar collagens) for collagen III. One role of FKBP65 seems to be to stabilise the collagen triple-helix and prevent premature association of collagen molecules. This is also the function of the collagen-specific chaperone HSP47, for which the crystal structure in complex with a triple-helical collagen peptide was recently determined (Widmer et al. 2012; Fig. 14.3d). The structure shows two molecules of HSP47 interacting with one collagen triple helix. Interactions involve an essential arginine residue in a Pro-Arg-Gly triplet of the helix and a conserved aspartate in HSP47, such that binding of two molecules of HSP47 to the leading and trailing strands of the helix precludes binding to the middle strand. Recent genetic data suggest that FKBP65 and HSP47 themselves interact during the control of procollagen biosynthesis (Duran et al. 2015b). FKBP65 is also essential for the conversion of lysyl hydroxylase 2 into its dimeric, active form (Gjaltema et al. 2016).
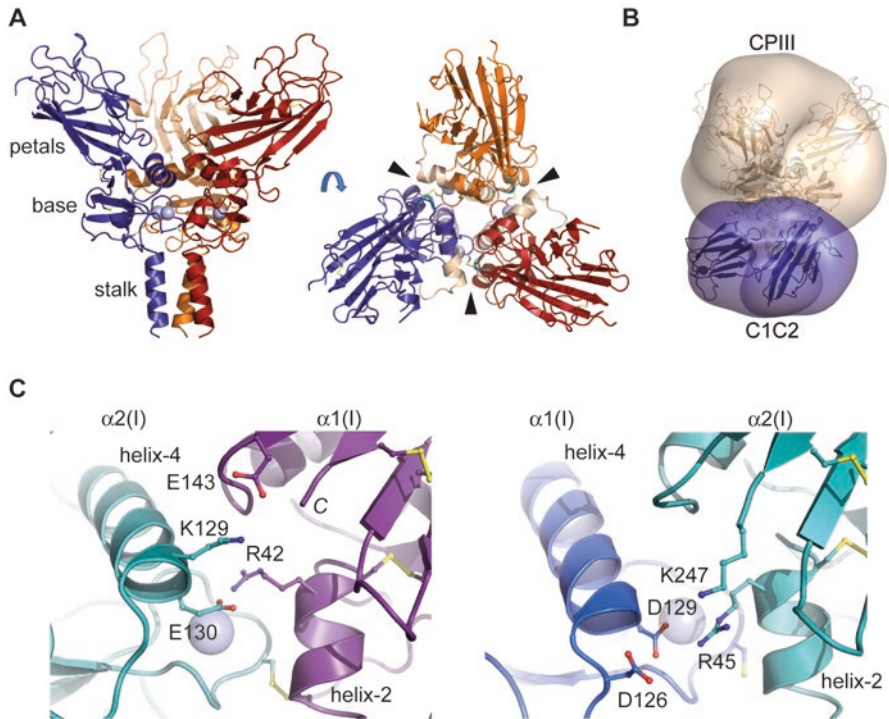
Fig. 14.5 Structure and interactions of procollagen C-propeptide trimers. (**a**) C-propeptide trimer from procollagen III (CPIII) (Bourhis et al. 2012). *Left*: side view. *Right*: top view showing the long (*cream colour*) and short (*blue/green*) stretches of the CRS interacting at each inter-chain interface (*arrows*). (**b**) Low resolution structure showing binding of the active C1C2 fragment of PCPE-1 to the stalk/base region of CPIII (Bourhis et al. 2013). (**c**) Interactions between the α2 chain (in turquoise) and the two neighbouring α1 chains (in *magenta* and *blue*) in the C-propeptide trimer of heterotrimeric procollagen I (Sharma et al. 2016)

14.7.3 Fibrillar Procollagen C-Propeptide Trimer

A major advance in recent years has been the first determination of the three-dimensional structure of a C-propeptide trimer (COLFI domain, Fig. 14.1), from procollagen III (Bourhis et al. 2012). COLFI domains are an essential component of fibrillar procollagens since the dawn of the metazoa (Exposito et al. 2010). Their role is to gather together the three polypeptide chains that make up each procollagen molecule, within the lumen of the rough endoplasmic reticulum, in order to initiate trimerization and folding which proceeds in a zipper-like manner towards the N-terminal end. This process is very specific, both in order to avoid the production of chimeric molecules containing different genetic types and to ensure the correct trimerization of chains in homotrimeric (collagen types II, III) and heterotrimeric (collagens I, V, XI) forms. While non-fibrillar collagens use other types of

trimerization domains, the structures of which are known (Boudko et al. 2012), those of the fibrillar procollagen COLFI domains have remained relatively elusive. In general, the sequences of the human COLFI domains (approx. 250 residues) are highly conserved (60 % identity). They have, however, a short and highly variable discontinuous sequence (15 residues) called the chain recognition sequence (CRS) that was shown almost 20 years ago to specify trimerization (Lees et al. 1997). For example, replacement of the CRS in the COLFI domain of the α2 chain of procollagen I, which does not trimerize on its own, by the CRS from procollagen III, leads to the formation of α2 chain trimers. This shows that the CRS contains the information required for chain association. The structure of the COLFI trimer from procollagen III has the shape of a flower, with a stalk, a base and three petals (Fig. 14.5a). When viewed from the top, the long stretch (12 residues) of the CRS on one chain is aligned with the short stretch (3 residues) in a neighbouring chain, thus showing the structural basis of chain recognition by the CRS.

Unlike procollagen III, an obligate homotrimer, procollagen I is normally an $[\alpha 1(I)]_2 \alpha 2(I)$ heterotrimer (Table 14.2). However, in embryonic tissues as well as in several diseases (cancer, fibrosis, genetic disorders), small amounts of the homotrimeric form $[\alpha 1(I)]_3$ are found. Until very recently, the molecular explanation for why procollagen I forms homotrimers and heterotrimers has been unknown. This problem has now been resolved by the elucidation of the structures of the homo- and heterotrimeric COLFI trimers of procollagen I (Sharma et al. 2016). Close inspection of the interactions at each of the two interfaces involving the α2(I) chain shows that the sequence of this chain has specifically evolved to permit interactions with α1(I) chains on either side whilst preventing the formation of an α2(I) chain homotrimer (Fig. 14.5c).

14.7.4 Proteolytic Processing of Procollagen

Proteolytic processing is another area where there has been significant progress in recent years. This involves the removal of the large N- and C-terminal propeptides by specific proteinases, normally ADAMTS2 and BMP-1, respectively (Bekhouche and Colige 2015; Vadon-Le Goff et al. 2015), resulting in the mature collagen molecule (Fig. 14.1). However, very recent observations have shown that ADAMTS2 also cleaves the C-propeptides from procollagen III, at a site N-terminal to the BMP-1 site, which appears to be a feature specific to this collagen type (Bekhouche et al. 2016). Concerning BMP-1, though it is not involved in N-terminal processing of the major fibrillar procollagens (types I, II and III), this proteinase is also involved in N-terminal processing of procollagen V (in addition to ADAMTS2) where both BMP-1 and furin are involved in C-terminal processing. To add to the complexity, recently it has also been shown that in skin, meprin proteinases cleave both the N- and C-propeptides of procollagen I, at sites close to those of ADAMTS2 and BMP-1 (Broder et al. 2013). Like the BMP-1/tolloid-like proteinases (BTPs), meprins are

members of the astacin family of zinc metalloproteinases, itself a member of the metzincin clan that includes the better known matrix metalloproteinases (MMPs). Thus, while originally thought to be relatively simple, there are now several proteinases that have been shown to be involved in procollagen processing *in vivo*. To further add to the complexity, the kinetics of BMP-1 cleavage of the C-propeptides of the fibrillar procollagens are increased (at least 10-fold) by another extracellular matrix protein, procollagen C-proteinase enhancer-1 (PCPE-1) (Bourhis et al. 2013). The mechanism of action of PCPE-1 involves specific binding of one molecule to the stalk and base region of the procollagen C-propeptide trimer (Fig. 14.5b). However, how this leads to enhancement of BMP-1 activity remains unknown.
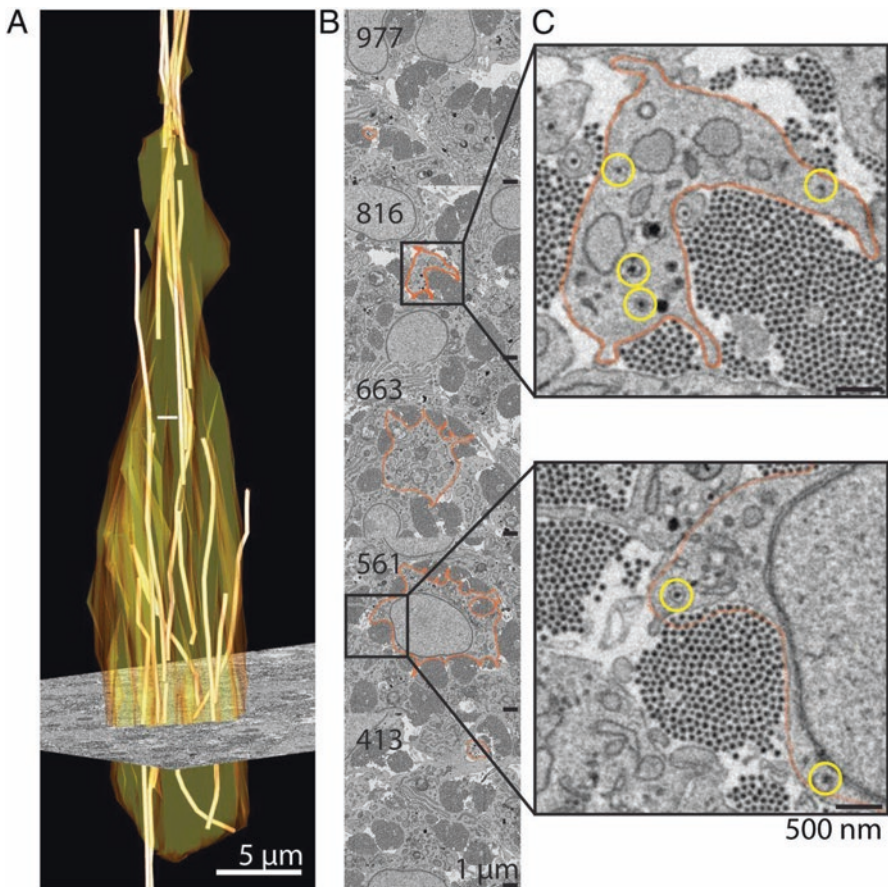
Fig. 14.6 Serial block face-scanning electron microscopy of collagen fibrils in a $30 \times 30 \times 100$ μm block of 14-day chick embryo tendon. (**a**) 3D reconstruction of a single cell (*outlined* in *orange*) showing intra-and extracellular collagen fibrils. (**b**) Images of selected areas from 5 of the 1000 cuts used to build up the entire structure. Plasma membranes outlined in orange. (**c**) Zooms of outlined areas in B. *Yellow circles* show collagen fibrils in cytoplasmic compartments. From (Kalson et al. 2013)

14.8   Fibril Assembly

It is now clear that the mechanisms involved in the assembly of collagen fibrils *in vivo* are far more complex than appeared in earlier *in vitro* observations (Kadler et al. 1996). An important advance has been the introduction of serial block face-scanning electron microscopy (SBF-SEM), which permits the three-dimensional reconstruction of collagen fibril organization in large tissue blocks (length 100 μm) to a resolution of 10 nm (Kalson et al. 2013) (Fig. 14.6). In embryonic chick tendon, extracellular collagens fibrils (~30 nm diameter) form parallel bundles in which the average fibril length is ~600 μm. In addition, fibrils are present in plasma membrane-bound compartments at the cell surface, called fibripositors, which either protrude outwards into the extracellular matrix or are recessed into the cell interior. There are also a small number of fibrils that are entirely enclosed by membrane, in so-called fibricarriers. The total number of membrane-bound compartments decreases when collagen synthesis is inhibited, suggesting that they are intermediates in the early stages of fibrillogenesis, a process which occurs in close proximity to the plasma membrane. Also, inhibition of myosin II activity eliminates fibricarriers suggesting that these are formed by active pulling on fibripositors by the cell.

Extracellular, small diameter (35 nm) collagen fibril bundles are also seen in embryonic mouse tendon (Kalson et al. 2015). Interestingly, while both the number of fibrils per bundle and the average fibril diameter increase up to birth, from birth to 6 weeks the number of fibrils per bundle remains constant while the average fibril diameter increases over threefold. Meanwhile, cells remain connected by cell-cell contacts with the number of bundles per cell remaining constant throughout. This suggests that up until birth there is both new fibril formation and fibril growth, the number of fibril bundles being determined by the cells, while post-birth, existing fibrils continue to increase in diameter. Interestingly, there is no change in fibril diameter with increasing distance from the cell. This argues against diffusion-limited accretion as a model for fibril growth in favour of interface-controlled growth at the fibril surface. Finally, it has recently been shown that in the absence of matrix metalloproteinase-14 (MMP-14), newly formed fibrils are not released from fibripositors, thus leading to fewer collagen fibrils and thinner tendons (Taylor et al. 2015). While MMP-14 is a collagenase, this effect does not require collagen cleavage but is thought to involve cleavage of fibronectin, a known nucleator of collagen fibril formation (Kadler et al. 2008).

14.9   Fibril Structure

Ten years ago, the first three-dimensional structure showing individual molecules within native collagen fibrils was described (Orgel et al. 2006). This structure was determined using fibre diffraction data from semi-crystalline rat tail tendon fibres, coupled with classical heavy atom isomorphous replacement, to build an electron density map with resolutions parallel and perpendicular to the fibre axis of 5 Å and

11 Å, respectively. As such, it did not allow individual amino acid residues to be identified, but did allow the paths of individual molecules to be traced (Fig. 14.4). In cross-section, molecules in the overlap region were arranged according to a quasi-hexagonal lattice with lateral dimensions of 40 Å and 27 Å, similar to those previously proposed (Hulmes and Miller 1979). Each unit cell contains five molecules in cross-section, numbered 1, 2, 3, 4 and 5 to denote the fact that neighbouring molecules are staggered by multiples of the axial D (670 Å) repeat. Prior to this structure, it was generally thought that molecules within each unit cell were arranged to form a so-called compressed microfibril (Piez and Trus 1981) running parallel to the fibril axis. In the new structure, the fibril axis is not parallel to the long (678 Å) c axis of the crystalline lattice and each molecule occupies five adjacent unit cells when moving along the a axis direction (Orgel et al. 2006). As a result, the microfibrils are not parallel to the fibril axis (Fig. 14.4) and also they are interdigitated with each other rather than being distinct structures. Furthermore, as predicted by Piez and Trus (1981), the tilt (about 5°) appears only in the comparatively tightly packed overlap region of the D period, molecules in the gap region being relatively less well organized and parallel to the fibril axis (Fig. 14.4). The validity of some details of this microfibrillar structure (Orgel et al. 2006) has been the subject of some controversy in the literature (Okuyama et al. 2009; Orgel 2009). There is agreement however that the paths of the molecules within the fibril are most likely correct, albeit that the resolution is insufficient to identify individual amino acid residues and in particular the order of the three chains within the triple helix.

While rat tail tendon collagen fibrils contain crystalline regions, they are usually circular in cross-section and therefore are not single crystals. Electron microscopy showed earlier that the characteristic ~38 Å lateral spacing in the crystalline lattice is oriented radially with respect to the fibril axis (Hulmes et al. 1985) and subsequently these data were used to model the overall fibril structure based on radially oriented crystalline regions separated by regions of disorder (Hulmes et al. 1995). Orgel et al. (2006) incorporated their new structural data into this model resulting in a similar concentric organization. An important feature of the Orgel structure, based on earlier data (Orgel et al. 2000, 2001), was the folded nature of the C-telopeptides of the α1(I) chains, such that the C-terminus of the molecule points towards the main triple helical region. This followed from heavy atom labelling data and also from the arrangement of cross-linking (hydroxy)lysines in adjacent 4D–staggered molecules in collagen fibrils. Comparison of surface bumps on collagen fibrils seen by scanning electron microscopy with models based on the relative bulkiness of the N- and C-telopeptides suggested that the C-telopeptides are located on the fibril surface (Perumal et al. 2008; Fig. 14.4) which is consistent with the radially oriented 38 Å spacing.

Collagen fibrils in rat tail tendons have relatively large diameters (up to 500 nm), while those in cornea and immature cartilage are relative narrow (~30 and 15 nm, respectively). Automated electron tomography of corneal fibrils (mainly type I collagen) also supports a concentric arrangement of tilted microfibrils (Holmes et al. 2001). Similarly, immature cartilage fibrils (mainly type II collagen) appear to be composed of a central core of 4 tilted microfibrils surrounded by an outer ring of 10

tilted microfibrils (Holmes and Kadler 2006). In both cases fibril diameters are strictly controlled by copolymerisation with minor fibrillar collagens, either type V collagen in cornea or type XI collagen in cartilage, both of which are characterized by delayed processing of N-terminal propeptides which are thought to coat the fibril surface thus preventing further accretion (see below). Interactions with small leucine rich proteoglycans (SLRPs) or non-fibrillar collagens are also known to control fibril diameter (Bruckner 2010; Chen et al. 2015; to be discussed later). The concentric arrangement in collagen fibrils implies that that there is something special about the fibril core. Evidence for a highly glycosylated fibril core in cross-sections of narrow collagen fibrils showing a pronounced helical twist has recently been obtained (Raspanti et al. 2011); this was not found however in larger fibrils (e.g. tendon) where the helical twist is smaller.

14.10 Fibril Cross-Linking and Degradation

Collagen fibrils are stabilized by the formation of covalent intermolecular cross-links (Yamauchi and Sricholpech 2012; Trackman 2016). These are initiated by the enzyme lysyl oxidase (LOX) which oxidatively deaminates lysine or hydroxylysine residues in the N- and C-terminal telopeptide regions. This then triggers spontaneous cross-link formation with (hydroxy)lysines in the triple-helical regions of adjacent molecules within the fibril (Fig. 14.4). These divalent cross-links can then undergo further maturation to form stable trivalent cross-links. Hydroxylysine residues can also be O-glycosylated by lysyl hydroxylases, and it has recently been shown that the type and extent of hydroxylysine glycosylation regulates cross-link maturation (Terajima et al. 2014). In addition, cross-linking can also be regulated by the small leucine rich proteoglycan fibromodulin, as shown by the increase in C-telopeptide collagen cross-linking seen in the tendons of fibromodulin-deficient mice (Kalamajski et al. 2014). In subsequent studies, the same group has shown that fibromodulin binds to the cross-linking regions in the collagen triple-helix that are adjacent to the telopeptide regions in neighbouring molecules within the fibril (Kalamajski et al. 2016; Fig. 14.4). These authors also showed that fibromodulin both binds to and increases the activity of LOX, suggesting that collagen, lysyl oxidase and fibromodulin might form a trimeric complex. It is well known that fibromodulin influences collagen fibrillogenesis (Kalamajski and Oldberg 2010). More recently however, it has also been shown that lysyl oxidase activity influences collagen fibril shape (Herchenhan et al. 2015) and even the complex lamellar organization of collagen fibrils seen in cornea (Wang et al. 2014).

Knowledge of the 3D structure of the collagen fibril has made it possible to model the many interactions with other components of the extracellular matrix that occur at the fibril surface (Orgel et al. 2011b). One such interaction is the digestion of collagen fibrils by matrix metalloproteinases (MMPs) (Fig. 14.3e; Manka et al. 2012). According to the Orgel structure, the MMP cleavage site, which occurs at 3/4 the length of the collagen molecule measured from the N-terminal end, is obstructed

by the C-terminal telopeptide (Fig. 14.4). Perumal et al. (2008) speculated that it would require prior proteolysis of the C-telopeptide in order to expose the MMP site. An alternative explanation is localised unfolding or breathing at hot spots along the fibril that facilitate MMP cleavage. Such hot spots were recently identified by tracking of individual fluorescently-tagged MMP1 molecules along collagen fibrils (Sarkar et al. 2012). Interestingly, this movement is directional, consistent with a "burnt bridge" mechanism in which cleavage of one collagen molecule exposes further cleavage sites on an adjacent molecule located C-terminal to the first (Saffarian et al. 2004). In addition, these hot spots were separated by a periodicity of ~1 μm, much more than the length of a single collagen molecule (0.3 μm). Very recently, this feature has been modelled in terms of periodic buckling (thus creating hot spots) as a means of relieving the internal strain brought about by lateral interactions (Dittmore et al. 2016).

14.11   Interactions of Fibrillar Collagens: Nucleators, Regulators and Organizers

14.11.1   Collagen Types V/XI Nucleate Heterotypic D-Period Collagen Fibrils

Collagen fibrils are in fact macromolecular alloys composed of several collagen types and other associated non-collagen proteins (Bruckner 2010). The heterotypic composition of the different fibrils depends on developmental stage and tissue, and it is becoming clear that referring to particular structures as "type I collagen fibril" or "type II collagen fibril" is an over-simplification (Bruckner 2010). Tendons, for the most part, are composed of type I collagen, but between 5 % and 10 % is type III collagen. However, during embryonic development type III is predominant. Skin collagen fibrils are also made of collagens I and III, cartilage collagen fibrils contain collagens II and XI, and corneal collagen fibrils contain collagens I and V (Ricard-Blum 2011; Mienaltowski and Birk 2014). The organization of these heterotypic fibrils can be complex. Cartilage fibrils are mainly collagen II but require collagen XI for nucleation and then for regulation of lateral growth (Blaschke et al. 2000; Kadler et al. 2008). In general, minor fibrillar collagens V and XI co-assemble with major fibrillar types I, II and III, and then other non-fibrillar collagen types localize to the fibril surface (see below). The minor collagens retain part of their N-propeptides, which cannot be incorporated into the lateral packing of collagen molecules in the interior of the fibril and must localise to the fibril surface through the gap regions. Through this mechanism the partially retained N-propeptides have a regulatory effect on fibril diameter (Birk 2001; Mienaltowski and Birk 2014).

14.11.2   *Collagen Fibril Regulators on the Fibril Surface*

There are two main groups of molecules that are associated with D-periodic fibrils but which neither form part of the fibrils themselves nor do they form fibrils on their own. They can be classified in the broad category of collagen fibril regulators (Kadler et al. 2008). One group is the family of FACITs, the Fibril Associated Collagens with Interrupted Triple helices. These are multi-domain collagen molecules with several COL and NC domains that are located on the surface of the collagen fibrils. Type IX collagen was the first FACIT to be discovered, cross-linked to type II collagen molecules on cartilage fibrils (Olsen 1997). The family grew later to include types XII and XIV associated with fibrils containing type I collagen in musculo-skeletal connective tissues (Chiquet et al. 2014), and then to the largely uncharacterized types XVI and XIX-XXII (Kadler et al. 2007; Ricard-Blum 2011; Mienaltowski and Birk 2014). Collagen types IX and XII have covalently attached glycosaminoglycan (GAG) chains and can therefore be considered as proteoglycans (Olsen 1997; Kadler et al. 2007; Chiquet et al. 2014). FACITs regulate collagen fibrils by affecting their surface properties (an analogy often used is that they "lubricate" the fibrils). Generally the COL domains of FACITs are short and with interruptions. The C-terminal region of the FACITs (domains NC1 and COL1) is thought to interact with the fibril surface, whereas the large globular domains at the N-terminus project outwards.

The second group of collagen fibril regulators are the family of Small Leucine Rich repeat Proteoglycans (SLRPs) (Chen and Birk 2013; Chen et al. 2015). SLRPs have a repetitive structure that is completely different to that of collagen. Their core proteins are made of tandem leucine-rich repeats (LRRs), a structural motif characterized by a relatively high proportion of leucine and other hydrophobic residues conserved in a repeating pattern. This particular structure generates curved solenoid LRR domains with an exposed β-sheet in the concave side that is suitable for protein-protein interactions (Bella et al. 2008). The core proteins of SLRPs have conserved cysteine clusters at both ends of the LRR domain, and are thought to interact with several collagen triple-helical molecules filling the concave space (Scott et al. 2004; McEwan et al. 2006; Bella et al. 2008; Orgel et al. 2009). SLRP protein cores are the regions mainly responsible for fibrillogenesis regulation (Chen and Birk 2013). The GAG chains extend away into the inter-fibrillar space where they are thought to interact with GAG chains from fibrils nearby, contributing to the regular lateral fibril organization seen in connective tissues such as cornea (Knupp et al. 2009).

Each family of collagen fibril regulators has its own specific expression pattern in time and space, including alternative splicing, and thus their interactions with cells and other components of the extracellular matrix are different and dynamic across tissues and at different developmental stages. This contributes to the diversity in structure and function of the different connective tissues (Kadler et al. 2007; Chen and Birk 2013; Chiquet et al. 2014; Chen et al. 2015).

14.11.3 Collagen Cell-Surface Receptors

Collagens are not just passive structural components of a molecular framework around which all the action occurs. Collagen biological function also involves several interactions with different cell-surface receptors and other extracellular matrix proteins. Interaction of collagens with specific cellular receptors triggers signalling events that regulate cell behaviour events such as adhesion, migration or proliferation. There are three main collagen receptor families: collagen-binding integrins, collagen-binding immune receptors and discoidin domain receptors (DDR1 and DDR2) (Heino 2007; Leitinger 2011; Hamaia and Farndale 2014; Heino 2014). Integrins are heterodimeric transmembrane glycoprotein receptors made of one α subunit and one β subunit. In humans there are eighteen and eight genes coding for α and β chains respectively, which associate in a number of combinations. Four of them are the collagen binding integrins α1β1, α2β1, α10β1 and α11β1, with different tissue distributions (Heino 2007, 2014; Hamaia and Farndale 2014; Hohenester 2014). The DDRs are kinase receptors that bind the major fibrillar collagen types I-III and regulate cell proliferation, differentiation and matrix modulation (Heino 2007; Leitinger and Hohenester 2007). Glycoprotein VI is an important platelet collagen receptor that regulates platelet function during thrombosis. Other immune-related receptors are leukocyte-associated immunoglobulin-like receptor-1 (LAIR-1), glycoprotein IV and OSCAR (Zhou et al. 2016). Finally, cell-surface proteoglycans such as syndecan, glypican or some isoforms of CD44 appear to bind collagen through their GAG chains (Heino 2007).

The triple-helical structure of the COL domains allows an essentially linear mapping between collagen sequence and the three-dimensional structure of molecular recognition motifs. These motifs must also be compatible with the repetitive (Gly-Xaa-Yaa)$_n$ sequence of the COL domains (Bella 2016). Libraries of peptides have been used to investigate specific collagen recognition sequences with different receptors and matrix molecules (Farndale et al. 2008; Konitsiotis et al. 2008; Hamaia et al. 2012; Manka et al. 2012; Hamaia and Farndale 2014). Thus, fibrillar collagens bind integrins through GXOGER and similar sequences, where X is often a hydrophobic residue and the R residue appears to have some variability (Hamaia and Farndale 2014), and DDR receptors recognize a GVMGFO sequence (Farndale et al. 2008; Konitsiotis et al. 2008). The structural details of triple-helical collagen recognition by other proteins have been clarified by crystal structure determinations (Fig. 14.3) (Bella 2016).

Because of the linear nature of the receptor-binding motifs, it is possible to map collagen-binding sites on the sequences of COL domains (interactome maps) and also on approximate 2D and 3D models of collagen fibrils based on fibre X-ray diffraction (Fig. 14.4; Orgel et al. 2006, 2011a, b; Sweeney et al. 2008). These maps help in identifying functional domains along the collagen sequences and the impact of known collagen mutations, proteolytic degradation or structural trauma on major ligand-binding sites.

14.12   Concluding Remarks

While the general structural features of fibrillar collagens have been known for some time, the field has recently grown in interest as new biochemical and functional information has emerged. Collagen model peptides have revealed fine details of the structure and stability of the triple helix that are currently being exploited in the engineering of novel, artificial collagens with unusual properties. Crystal structures of procollagen C-propeptides have provided crucial structural information on the mechanisms of fibrillar collagen trimerization and chain selection. Furthermore, molecular details of triple helical collagen recognition by receptors and other collagen-binding proteins are constantly being discovered through the use of peptide libraries and the determination of crystal structures of complexes with collagen peptides. All this information will be essential for our understanding of an increasingly complex scenario in which the particular composition and structure of fibrillar collagens appears to change temporally and spatially for every connective tissue and developmental stage. Ultimately, the combination of structural and functional information will be needed to understand the unique physiological characteristics of each connective tissue and the pathological consequences of connective tissue disorders.

References

Acevedo-Jake AM, Clements KA, Hartgerink JD (2016) Synthetic, register-specific, AAB heterotrimers to investigate single point glycine mutations in Osteogenesis imperfecta. Biomacromolecules 17(3):914–921. doi:10.1021/acs.biomac.5b01562

Anantharajan J, Koski MK, Kursula P, Hieta R, Bergmann U, Myllyharju J, Wierenga RK (2013) The structural motifs for substrate binding and dimerization of the alpha subunit of collagen prolyl 4-hydroxylase. Structure 21(12):2107–2118. doi:10.1016/j.str.2013.09.005

Bann JG, Peyton DH, Bachinger HP (2000) Sweet is stable: glycosylation stabilizes collagen. FEBS Lett 473(2):237–240

Bann JG, Bachinger HP, Peyton DH (2003) Role of carbohydrate in stabilizing the triple-helix in a model for a deep-sea hydrothermal vent worm collagen. Biochemistry 42(14):4042–4048. doi:10.1021/bi027050w

Bekhouche M, Colige A (2015) The procollagen N-proteinases ADAMTS2, 3 and 14 in pathophysiology. Matrix Biol 44-46:46–53. doi:10.1016/j.matbio.2015.04.001

Bekhouche M, Leduc C, Dupont L, Janssen L, Delolme F, Vadon-Le Goff S, Smargiasso N, Baiwir D, Mazzucchelli G, Zanella-Cleon I, Dubail J, De Pauw E, Nusgens B, Hulmes DJS, Moali C, Colige A (2016) Determination of the substrate repertoire of ADAMTS2, 3, and 14 significantly broadens their functions and identifies extracellular matrix organization and TGF-beta signaling as primary targets. FASEB J 30(5):1741–1756. doi:10.1096/fj.15-279869

Bella J (2010) A new method for describing the helical conformation of collagen: dependence of the triple helical twist on amino acid sequence. J Struct Biol 170(2):377–391. doi:10.1016/j.jsb.2010.02.003

Bella J (2014) A first census of collagen interruptions: collagen's own stutters and stammers. J Struct Biol 186(3):438–450. doi:10.1016/j.jsb.2014.03.022

Bella J (2016) Collagen structure: new tricks from a very old dog. Biochem J 473(8):1001–1025. doi:10.1042/BJ20151169

Bella J, Hindle KL, McEwan PA, Lovell SC (2008) The leucine-rich repeat structure. Cell Mol Life Sci 65(15):2307–2333. doi:10.1007/s00018-008-8019-0

Bender E, Silver FH, Hayashi K, Trelstad RL (1982) Type I collagen segment long spacing banding patterns. Evidence that the alpha 2 chain is in the reference or A position. J Biol Chem 257(16):9653–9657

Birk DE (2001) Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly. Micron 32(3):223–237

Blaschke UK, Eikenberry EF, Hulmes DJS, Galla HJ, Bruckner P (2000) Collagen XI nucleates self-assembly and limits lateral growth of cartilage fibrils. J Biol Chem 275(14):10370–10378

Boot-Handford RP, Tuckwell DS (2003) Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. Bioessays 25(2):142–151. doi:10.1002/bies.10230

Boot-Handford RP, Tuckwell DS, Plumb DA, Rock CF, Poulsom R (2003) A novel and highly conserved collagen (pro(alpha)1(XXVII)) with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. J Biol Chem 278(33):31067–31077. doi:10.1074/jbc.M212889200

Boudko SP, Engel J, Bachinger HP (2012) The crucial role of trimerization domains in collagen folding. Int J Biochem Cell Biol 44(1):21–32. doi:10.1016/j.biocel.2011.09.009

Bourhis JM, Mariano N, Zhao Y, Harlos K, Exposito JY, Jones EY, Moali C, Aghajari N, Hulmes DJS (2012) Structural basis of fibrillar collagen trimerization and related genetic disorders. Nat Struct Mol Biol 19(10):1031–1036. doi:10.1038/nsmb.2389

Bourhis JM, Vadon-Le Goff S, Afrache H, Mariano N, Kronenberg D, Thielens N, Moali C, Hulmes DJS (2013) Procollagen C-proteinase enhancer grasps the stalk of the C-propeptide trimer to boost collagen precursor maturation. Proc Natl Acad Sci U S A 110(16):6394–6399. doi:10.1073/pnas.1300480110

Broder C, Arnold P, Vadon-Le Goff S, Konerding MA, Bahr K, Muller S, Overall CM, Bond JS, Koudelka T, Tholey A, Hulmes DJS, Moali C, Becker-Pauly C (2013) Metalloproteases meprin alpha and meprin beta are C- and N-procollagen proteinases important for collagen assembly and tensile strength. Proc Natl Acad Sci U S A 110(35):14219–14224. doi:10.1073/pnas.1305464110

Brondijk TH, Bihan D, Farndale RW, Huizinga EG (2012) Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. Proc Natl Acad Sci U S A 109(14):5253–5258. doi:10.1073/pnas.1112388109

Bruckner P (2010) Suprastructures of extracellular matrices: paradigms of functions controlled by aggregates rather than molecules. Cell Tissue Res 339(1):7–18. doi:10.1007/s00441-009-0864-0

Cabral WA, Perdivara I, Weis M, Terajima M, Blissett AR, Chang W, Perosky JE, Makareeva EN, Mertz EL, Leikin S, Tomer KB, Kozloff KM, Eyre DR, Yamauchi M, Marini JC (2014) Abnormal type I collagen post-translational modification and crosslinking in a cyclophilin B KO mouse model of recessive osteogenesis imperfecta. PLoS Genet 10(6):e1004465. doi:10.1371/journal.pgen.1004465

Chen S, Birk DE (2013) The regulatory roles of small leucine-rich proteoglycans in extracellular matrix assembly. FEBS J 280(10):2120–2137. doi:10.1111/febs.12136

Chen S, Mienaltowski MJ, Birk DE (2015) Regulation of corneal stroma extracellular matrix assembly. Exp Eye Res 133:69–80. doi:10.1016/j.exer.2014.08.001

Chiquet M, Birk DE, Bonnemann CG, Koch M (2014) Collagen XII: protecting bone and muscle integrity by organizing collagen fibrils. Int J Biochem Cell Biol 53:51–54. doi:10.1016/j.biocel.2014.04.020

D'Arcangelo JG, Stahmer KR, Miller EA (2013) Vesicle-mediated export from the ER: COPII coat function and regulation. Biochim Biophys Acta 1833(11):2464–2472. doi:10.1016/j.bbamcr.2013.02.003

Daubenspeck JM, Zeng H, Chen P, Dong S, Steichen CT, Krishna NR, Pritchard DG, Turnbough CL Jr (2004) Novel oligosaccharide side chains of the collagen-like region of BclA, the major

glycoprotein of the *Bacillus anthracis* exosporium. J Biol Chem 279(30):30945–30953. doi:10.1074/jbc.M401613200

Dittmore A, Silver J, Sarkar SK, Marmer B, Goldberg GI, Neuman KC (2016) Internal strain drives spontaneous periodic buckling in collagen and regulates remodeling. Proc Natl Acad Sci U S A 113(30):8436–8441. doi:10.1073/pnas.1523228113

Dubey K, Kar K (2014) Type I collagen prevents amyloid aggregation of hen egg white lysozyme. Biochem Biophys Res Commun 448(4):480–484. doi:10.1016/j.bbrc.2014.04.135

Duran I, Csukasi F, Taylor SP, Krakow D, Becerra J, Bombarely A, Mari-Beffa M (2015a) Collagen duplicate genes of bone and cartilage participate during regeneration of zebrafish fin skeleton. Gene Expr Patterns 19(1–2):60–69. doi:10.1016/j.gep.2015.07.004

Duran I, Nevarez L, Sarukhanov A, Wu S, Lee K, Krejci P, Weis M, Eyre D, Krakow D, Cohn DH (2015b) HSP47 and FKBP65 cooperate in the synthesis of type I procollagen. Hum Mol Genet 24(7):1918–1928. doi:10.1093/hmg/ddu608

Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC (2000) Structural basis of collagen recognition by integrin α2β1. Cell 101(1):47–56. doi:10.1016/S0092-8674(00)80622-4

Exposito JY, Larroux C, Cluzel C, Valcourt U, Lethias C, Degnan BM (2008) Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. J Biol Chem 283(42):28226–28235. doi:10.1074/jbc.M804573200

Exposito JY, Valcourt U, Cluzel C, Lethias C (2010) The fibrillar collagen family. Int J Mol Sci 11(2):407–426. doi:10.3390/ijms11020407

Eyre DR, Weis MA, Wu JJ (2006) Articular cartilage collagen: an irreplaceable framework? Eur Cell Mater 12:57–63. doi:10.22203/eCM

Fallas JA, Dong J, Tao YJ, Hartgerink JD (2012) Structural insights into charge pair interactions in triple helical collagen-like proteins. J Biol Chem 287(11):8039–8047. doi:10.1074/jbc.M111.296574

Fang M, Jacob R, McDougal O, Oxford JT (2012) Minor fibrillar collagens, variable regions alternative splicing, intrinsic disorder, and tyrosine sulfation. Protein Cell 3(6):419–433. doi:10.1007/s13238-012-2917-5

Farndale RW, Lisman T, Bihan D, Hamaia S, Smerling CS, Pugh N, Konitsiotis A, Leitinger B, de Groot PG, Jarvis GE, Raynal N (2008) Cell-collagen interactions: the use of peptide Toolkits to investigate collagen-receptor interactions. Biochem Soc Trans 36(Pt 2):241–250. doi:10.1042/BST0360241

Fichard A, Kleman JP, Ruggiero F (1995) Another look at collagen V and XI molecules. Matrix Biol 14(7):515–531. doi:10.1016/S0945-053X(05)80001-0

Ghosh N, McKillop TJ, Jowitt TA, Howard M, Davies H, Holmes DF, Roberts IS, Bella J (2012) Collagen-like proteins in pathogenic *E. coli* strains. PLoS One 7(6):e37872. doi:10.1371/journal.pone.0037872

Gjaltema RA, van der Stoel MM, Boersema M, Bank RA (2016) Disentangling mechanisms involved in collagen pyridinoline cross-linking: the immunophilin FKBP65 is critical for dimerization of lysyl hydroxylase 2. Proc Natl Acad Sci U S A 113(26):7142–7147. doi:10.1073/pnas.1600074113

Gordon MK, Hahn RA (2010) Collagens. Cell Tissue Res 339(1):247–257. doi:10.1007/s00441-009-0844-4

Gorres KL, Raines RT (2010) Prolyl 4-hydroxylase. Crit Rev Biochem Mol Biol 45(2):106–124. doi:10.3109/10409231003627991

Hamaia S, Farndale RW (2014) Integrin recognition motifs in the human collagens. Adv Exp Med Biol 819:127–142. doi:10.1007/978-94-017-9153-3_9

Hamaia SW, Pugh N, Raynal N, Nemoz B, Stone R, Gullberg D, Bihan D, Farndale RW (2012) Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin α1β1 using collagen toolkits II and III. J Biol Chem 287(31):26019–26028. doi:10.1074/jbc.M112.353144

Heino J (2007) The collagen family members as cell adhesion proteins. Bioessays 29(10):1001–1010. doi:10.1002/bies.20636

Heino J (2014) Cellular signaling by collagen-binding integrins. Adv Exp Med Biol 819:143–155. doi:10.1007/978-94-017-9153-3_10

Herchenhan A, Uhlenbrock F, Eliasson P, Weis M, Eyre D, Kadler KE, Magnusson SP, Kjaer M (2015) Lysyl oxidase activity is required for ordered collagen fibrillogenesis by tendon cells. J Biol Chem 290(26):16440–16450. doi:10.1074/jbc.M115.641670

Hoffman GG, Branam AM, Huang G, Pelegri F, Cole WG, Wenstrup RM, Greenspan DS (2010) Characterization of the six zebrafish clade B fibrillar procollagen genes, with evidence for evolutionarily conserved alternative splicing within the pro-alpha1(V) C-propeptide. Matrix Biol 29(4):261–275. doi:10.1016/j.matbio.2010.01.006

Hofmann H, Fietzek PP, Kuhn K (1978) The role of polar and hydrophobic interactions for the molecular packing of type I collagen: a three-dimensional evaluation of the amino acid sequence. J Mol Biol 125(2):137–165. doi:10.1016/0022-2836(78)90342-X

Hohenester E (2014) Signalling complexes at the cell-matrix interface. Curr Opin Struct Biol 29:10–16. doi:10.1016/j.sbi.2014.08.009

Hohenester E, Sasaki T, Giudici C, Farndale RW, Bachinger HP (2008) Structural basis of sequence-specific collagen recognition by SPARC. Proc Natl Acad Sci U S A 105(47):18273–18277. doi:10.1073/pnas.0808452105

Holmes DF, Kadler KE (2006) The 10+4 microfibril structure of thin cartilage fibrils. Proc Natl Acad Sci U S A 103(46):17249–17254. doi:10.1073/pnas.0608417103

Holmes DF, Gilpin CJ, Baldock C, Ziese U, Koster AJ, Kadler KE (2001) Corneal collagen fibril structure in three dimensions: structural insights into fibril assembly, mechanical properties, and tissue organization. Proc Natl Acad Sci U S A 98(13):7307–7312. doi:10.1073/pnas.111150598

Hudson DM, Eyre DR (2013) Collagen prolyl 3-hydroxylation: a major role for a minor post-translational modification? Connect Tissue Res 54(4–5):245–251. doi:10.3109/03008207.2013.800867

Hudson DM, Werther R, Weis M, Wu JJ, Eyre DR (2014) Evolutionary origins of C-terminal (GPP)$_n$ 3-hydroxyproline formation in vertebrate tendon collagen. PLoS One 9(4):e93467. doi:10.1371/journal.pone.0093467

Hulmes DJS (2008a) Collagen diversity, synthesis and assembly. In: Fratzl P (ed) Collagen – structure and mechanics. Springer, New York, pp 15–47

Hulmes DJS (2008b) Vertebrate collagens – structures, functions and biomedical applications. In: Scheibel T (ed) Fibrous proteins. Landes Biosciences, Austin, pp 12–29

Hulmes DJS, Miller A (1979) Quasi-hexagonal molecular packing in collagen fibrils. Nature 282(5741):878–880

Hulmes DJS, Holmes DF, Cummings C (1985) Crystalline regions in collagen fibrils. J Mol Biol 184(3):473–477

Hulmes DJS, Wess TJ, Prockop DJ, Fratzl P (1995) Radial packing, order, and disorder in collagen fibrils. Biophys J 68(5):1661–1670. doi:10.1016/S0006-3495(95)80391-7

Huxley-Jones J, Robertson DL, Boot-Handford RP (2007) On the origins of the extracellular matrix in vertebrates. Matrix Biol 26(1):2–11. doi:10.1016/j.matbio.2006.09.008

Ishikawa Y, Bachinger HP (2013) A molecular ensemble in the rER for procollagen maturation. Biochim Biophys Acta 1833(11):2479–2491. doi:10.1016/j.bbamcr.2013.04.008

Ishikawa Y, Boudko S, Bachinger HP (2015) Ziploc-ing the structure: triple helix formation is coordinated by rough endoplasmic reticulum resident PPIases. Biochim Biophys Acta 1850(10):1983–1993. doi:10.1016/j.bbagen.2014.12.024

Kadler KE, Holmes DF, Trotter JA, Chapman JA (1996) Collagen fibril formation. Biochem J 316(Pt 1):1–11

Kadler KE, Baldock C, Bella J, Boot-Handford RP (2007) Collagens at a glance. J Cell Sci 120(Pt 12):1955–1958. doi:10.1242/jcs.03453

Kadler KE, Hill A, Canty-Laird EG (2008) Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. Curr Opin Cell Biol 20(5):495–501. doi:10.1016/j.ceb.2008.06.008

Kalamajski S, Oldberg A (2010) The role of small leucine-rich proteoglycans in collagen fibrillogenesis. Matrix Biol 29(4):248–253. doi:10.1016/j.matbio.2010.01.001

Kalamajski S, Liu C, Tillgren V, Rubin K, Oldberg A, Rai J, Weis M, Eyre DR (2014) Increased C-telopeptide cross-linking of tendon type I collagen in fibromodulin-deficient mice. J Biol Chem 289(27):18873–18879. doi:10.1074/jbc.M114.572941

Kalamajski S, Bihan D, Bonna A, Rubin K, Farndale RW (2016) Fibromodulin interacts with collagen cross-linking sites and activates lysyl oxidase. J Biol Chem 291(15):7951–7960. doi:10.1074/jbc.M115.693408

Kalson NS, Starborg T, Lu Y, Mironov A, Humphries SM, Holmes DF, Kadler KE (2013) Nonmuscle myosin II powered transport of newly formed collagen fibrils at the plasma membrane. Proc Natl Acad Sci U S A 110(49):E4743–E4752. doi:10.1073/pnas.1314348110

Kalson NS, Lu Y, Taylor SH, Starborg T, Holmes DF, Kadler KE (2015) A structure-based extracellular matrix expansion mechanism of fibrous tissue growth. eLife 4:e05958. doi:10.7554/eLife.05958

Keen AN, Fenna AJ, McConnell JC, Sherratt MJ, Gardner P, Shiels HA (2016) The dynamic nature of hypertrophic and fibrotic remodeling of the fish ventricle. Front Physiol 6:427. doi:10.3389/fphys.2015.00427

Knupp C, Pinali C, Lewis PN, Parfitt GJ, Young RD, Meek KM, Quantock AJ (2009) The architecture of the cornea and structural basis of its transparency. Adv Protein Chem Struct Biol 78:25–49. doi:10.1016/S1876-1623(08)78002-7

Koch M, Laub F, Zhou P, Hahn RA, Tanaka S, Burgeson RE, Gerecke DR, Ramirez F, Gordon MK (2003) Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone. J Biol Chem 278(44):43236–43244. doi:10.1074/jbc.M302112200

Konitsiotis AD, Raynal N, Bihan D, Hohenester E, Farndale RW, Leitinger B (2008) Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. J Biol Chem 283(11):6861–6868. doi:10.1074/jbc.M709290200

Kramer RZ, Bella J, Brodsky B, Berman HM (2001) The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. J Mol Biol 311(1):131–147. doi:10.1006/jmbi.2001.4849

Lees JF, Tasab M, Bulleid NJ (1997) Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. EMBO J 16(5):908–916. doi:10.1093/emboj/16.5.908

Leikina E, Mertts MV, Kuznetsova N, Leikin S (2002) Type I collagen is thermally unstable at body temperature. Proc Natl Acad Sci U S A 99(3):1314–1318. doi:10.1073/pnas.032307099

Leitinger B (2011) Transmembrane collagen receptors. Annu Rev Cell Dev Biol 27:265–290. doi:10.1146/annurev-cellbio-092910-154013

Leitinger B, Hohenester E (2007) Mammalian collagen receptors. Matrix Biol 26(3):146–155. doi:10.1016/j.matbio.2006.10.007

Mak KM, Png CY, Lee DJ (2016) Type V collagen in health, disease, and fibrosis. Anat Rec (Hoboken) 299(5):613–629. doi:10.1002/ar.23330

Manka SW, Carafoli F, Visse R, Bihan D, Raynal N, Farndale RW, Murphy G, Enghild JJ, Hohenester E, Nagase H (2012) Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. Proc Natl Acad Sci U S A 109(31):12461–12466. doi:10.1073/pnas.1204991109

McAlinden A (2014) Alternative splicing of type II procollagen: IIB or not IIB? Connect Tissue Res 55(3):165–176. doi:10.3109/03008207.2014.908860

McEwan PA, Scott PG, Bishop PN, Bella J (2006) Structural correlations in the family of small leucine-rich repeat proteins and proteoglycans. J Struct Biol 155(2):294–305. doi:10.1016/j.jsb.2006.01.016

Mienaltowski MJ, Birk DE (2014) Structure, physiology, and biochemistry of collagens. Adv Exp Med Biol 802:5–29. doi:10.1007/978-94-007-7893-1_2

Myllyharju J (2005) Intracellular post-translational modifications of collagens. Top Curr Chem 247:115–247. doi:10.1007/b103821

Myllyharju J (2008) Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. Ann Med 40(6):402–417. doi:10.1080/07853890801986594

Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet 20(1):33–43. doi:10.1016/j.tig.2003.11.004

Nishi Y, Uchiyama S, Doi M, Nishiuchi Y, Nakazawa T, Ohkubo T, Kobayashi Y (2005) Different effects of 4-hydroxyproline and 4-fluoroproline on the stability of collagen triple helix. Biochemistry 44(16):6034–6042. doi:10.1021/bi047887m

Okuyama K, Bachinger HP, Mizuno K, Boudko S, Engel J, Berisio R, Vitagliano L (2009) Re: microfibrillar structure of type I collagen in situ. Acta Crystallogr D Biol Crystallogr 65(Pt 9):1007–1008. doi:10.1107/S0907444909023051author reply 1009-1010

Olsen BR (1997) Collagen IX. Int J Biochem Cell Biol 29(4):555–558

Orgel JPRO (2009) On the packing structure of collagen: response to Okuyama et al.'s comment on Microfibrillar Structure of Type I Collagens in situ. Acta Cryst D65:1009–1010

Orgel JPRO, Wess TJ, Miller A (2000) The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen. Structure 8(2):137–142

Orgel JPRO, Miller A, Irving TC, Fischetti RF, Hammersley AP, Wess TJ (2001) The in situ supermolecular structure of type I collagen. Structure 9(11):1061–1069

Orgel JPRO, Irving TC, Miller A, Wess TJ (2006) Microfibrillar structure of type I collagen in situ. Proc Natl Acad Sci U S A 103(24):9001–9005. doi:10.1073/pnas.0502718103

Orgel JPRO, Eid A, Antipova O, Bella J, Scott JE (2009) Decorin core protein (decoron) shape complements collagen fibril surface structure and mediates its binding. PLoS One 4(9):e7028. doi:10.1371/journal.pone.0007028

Orgel JPRO, Antipova O, Sagi I, Bitler A, Qiu D, Wang R, Xu Y, San Antonio JD (2011a) Collagen fibril surface displays a constellation of sites capable of promoting fibril assembly, stability, and hemostasis. Connect Tissue Res 52(1):18–24. doi:10.3109/03008207.2010.511354

Orgel JPRO, San Antonio JD, Antipova O (2011b) Molecular and structural mapping of collagen fibril interactions. Connect Tissue Res 52(1):2–17. doi:10.3109/03008207.2010.511353

Orgel JPRO, Persikov AV, Antipova O (2014) Variation in the helical structure of native collagen. PLoS One 9(2):e89519. doi:10.1371/journal.pone.0089519

Page AP, Johnstone IL (2007) The cuticle. WormBook:1–15. doi:10.1895/wormbook.1.138.1

Parmar AS, Nunes AM, Baum J, Brodsky B (2012) A peptide study of the relationship between the collagen triple-helix and amyloid. Biopolymers 97(10):795–806. doi:10.1002/bip.22070

Persikov AV, Brodsky B (2002) Unstable molecules form stable tissues. Proc Natl Acad Sci U S A 99(3):1101–1103. doi:10.1073/pnas.042707899

Persikov AV, Ramshaw JA, Brodsky B (2005) Prediction of collagen stability from amino acid sequence. J Biol Chem 280(19):19343–19349. doi:10.1074/jbc.M501657200

Perumal S, Antipova O, Orgel JPRO (2008) Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. Proc Natl Acad Sci U S A 105(8):2824–2829. doi:10.1073/pnas.0710588105

Piez KA, Trus BL (1981) A new model for packing of type-I collagen molecules in the native fibril. Biosci Rep 1(10):801–810

Plumb DA, Dhir V, Mironov A, Ferrara L, Poulsom R, Kadler KE, Thornton DJ, Briggs MD, Boot-Handford RP (2007) Collagen XXVII is developmentally regulated and forms thin fibrillar structures distinct from those of classical vertebrate fibrillar collagens. J Biol Chem 282(17):12791–12795. doi:10.1074/jbc.C700021200

Raspanti M, Reguzzoni M, Protasoni M, Martini D (2011) Evidence of a discrete axial structure in unimodal collagen fibrils. Biomacromolecules 12 (12):4344–4347. doi:10.1021/bm201314e

Ricard-Blum S (2011) The collagen family. Cold Spring Harb Perspect Biol 3(1):a004978. doi:10.1101/cshperspect.a004978

Rosset EM, Bradshaw AD (2016) SPARC/osteonectin in mineralized tissue. Matrix Biol 52-54:78–87. doi:10.1016/j.matbio.2016.02.001

Saffarian S, Collier IE, Marmer BL, Elson EL, Goldberg G (2004) Interstitial collagenase is a Brownian ratchet driven by proteolysis of collagen. Science 306(5693):108–111. doi:10.1126/science.1099179

Sarkar SK, Marmer B, Goldberg G, Neuman KC (2012) Single-molecule tracking of collagenase on native type I collagen fibrils reveals degradation mechanism. Curr Biol 22(12):1047–1056. doi:10.1016/j.cub.2012.04.012

Schumacher MA, Mizuno K, Bachinger HP (2006) The crystal structure of a collagen-like polypeptide with 3(S)-hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. J Biol Chem 281(37):27566–27574. doi:10.1074/jbc.M602797200

Scott PG, McEwan PA, Dodd CM, Bergmann EM, Bishop PN, Bella J (2004) Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. Proc Natl Acad Sci U S A 101(44):15633–15638. doi:10.1073/pnas.0402976101

Sharma U, Carrique L, Vadon-Le Goff S, Mariano N, Georges RN, Delolme F, Koivunen P, Myllyharju J, Moali C, Aghajari N, Hulmes DJS (2016) Structural basis of homo- and heterotrimerization of collagen I. Nature Commun, in press

Shoulders MD, Raines RT (2009) Collagen structure and stability. Annu Rev Biochem 78:929–958. doi:10.1146/annurev.biochem.77.032207.120833

Shoulders MD, Satyshur KA, Forest KT, Raines RT (2010) Stereoelectronic and steric effects in side chains preorganize a protein main chain. Proc Natl Acad Sci U S A 107(2):559–564. doi:10.1073/pnas.0909592107

Smith SM, Birk DE (2012) Focus on molecules: collagens V and XI. Exp Eye Res 98:105–106. doi:10.1016/j.exer.2010.08.003

Sun M, Chen S, Adams SM, Florer JB, Liu H, Kao WW, Wenstrup RJ, Birk DE (2011) Collagen V is a dominant regulator of collagen fibrillogenesis: dysfunctional regulation of structure and function in a corneal-stroma-specific Col5a1-null mouse model. J Cell Sci 124(Pt 23):4096–4105. doi:10.1242/jcs.091363

Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, Forlino A, Cabral WA, Barnes AM, Marini JC, San Antonio JD (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J Biol Chem 283(30):21187–21197. doi:10.1074/jbc.M709319200

Taga Y, Kusubata M, Ogawa-Goto K, Hattori S (2016) Developmental stage-dependent regulation of prolyl 3-hydroxylation in tendon type I collagen. J Biol Chem 291(2):837–847. doi:10.1074/jbc.M115.686105

Taylor SH, Yeung CY, Kalson NS, Lu Y, Zigrino P, Starborg T, Warwood S, Holmes DF, Canty-Laird EG, Mauch C, Kadler KE (2015) Matrix metalloproteinase 14 is required for fibrous tissue expansion. Elife 4:e09345. doi:10.7554/eLife.09345

Terajima M, Perdivara I, Sricholpech M, Deguchi Y, Pleshko N, Tomer KB, Yamauchi M (2014) Glycosylation and cross-linking in bone type I collagen. J Biol Chem 289(33):22636–22647. doi:10.1074/jbc.M113.528513

Thom JR, Morris NP (1991) Biosynthesis and proteolytic processing of type XI collagen in embryonic chick sterna. J Biol Chem 266(11):7262–7269

Trackman PC (2016) Enzymatic and non-enzymatic functions of the lysyl oxidase family in bone. Matrix Biol 52-54:7–18. doi:10.1016/j.matbio.2016.01.001

Vadon-Le Goff S, Hulmes DJS, Moali C (2015) BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. Matrix Biol 44-46:14–23. doi:10.1016/j.matbio.2015.02.006

Vitagliano L, Berisio R, Mazzarella L, Zagari A (2001) Structural bases of collagen stabilization induced by proline hydroxylation. Biopolymers 58(5):459–464. doi:10.1002/1097-0282(20010415)58:5<459::AID-BIP1021>3.0.CO;2-V

Vranka JA, Pokidysheva E, Hayashi L, Zientek K, Mizuno K, Ishikawa Y, Maddox K, Tufa S, Keene DR, Klein R, Bachinger HP (2010) Prolyl 3-hydroxylase 1 null mice display abnormalities in fibrillar collagen-rich tissues such as tendons, skin, and bones. J Biol Chem 285(22):17253–17262. doi:10.1074/jbc.M110.102228

Wang L, Uhlig PC, Eikenberry EF, Robenek H, Bruckner P, Hansen U (2014) Lateral growth limitation of corneal fibrils and their lamellar stacking depend on covalent collagen cross-linking by transglutaminase-2 and lysyl oxidases, respectively. J Biol Chem 289(2):921–929. doi:10.1074/jbc.M113.496364

Weis MA, Hudson DM, Kim L, Scott M, Wu JJ, Eyre DR (2010) Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. J Biol Chem 285(4):2580–2590. doi:10.1074/jbc.M109.068726

Widmer C, Gebauer JM, Brunstein E, Rosenbaum S, Zaucke F, Drogemuller C, Leeb T, Baumann U (2012) Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. Proc Natl Acad Sci U S A 109(33):13243–13247. doi:10.1073/pnas.1208072109

Yamauchi M, Sricholpech M (2012) Lysine post-translational modifications of collagen. Essays Biochem 52:113–133. doi:10.1042/bse0520113

Zhou L, Hinerman JM, Blaszczyk M, Miller JL, Conrady DG, Barrow AD, Chirgadze DY, Bihan D, Farndale RW, Herr AB (2016) Structural basis for collagen recognition by the immune receptor OSCAR. Blood 127(5):529–537. doi:10.1182/blood-2015-08-667055

Chapter 15
Recombinant Structural Proteins and Their Use in Future Materials

Tara D. Sutherland, Trevor D. Rapson, Mickey G. Huson, and Jeffrey S. Church

Contents

Abstract Recombinant proteins are polymers that offer the materials engineer absolute control over chain length and composition: key attributes required for design of advanced polymeric materials. Through this control, these polymers can be encoded to contain information that enables them to respond as the environment changes. However, despite their promise, protein-based materials are under-

T.D. Sutherland (✉) • T.D. Rapson
CSIRO, Black Mountain, 2601 Canberra, ACT, Australia
e-mail: Tara.Sutherland@CSIRO.AU; Trevor.Rapson@CSIRO.AU

M.G. Huson • J.S. Church
CSIRO, Waurn Ponds, 3216 Geelong, VIC, Australia
e-mail: Mickey.Huson@CSIRO.AU; Jeff.Church@CSIRO.AU

represented in materials science. In this chapter we investigate why this is and describe recent efforts to address this. We discuss constraints limiting rational design of structural proteins for advanced materials; advantages and disadvantages of different recombinant expression platforms; and, methods to fabricate proteins into solid-state materials. Finally, we describe the silk proteins used in our laboratory as templates for information-containing polymers.

Keywords Recombinant protein • Advanced materials • Silk • Collagen • Resilin • Elastin

15.1   Introduction

The oil age revolutionised the use of polymers in materials science, with polymeric materials now playing a critical role throughout the modern world. Development of the next generation of materials, for sophisticated applications such as regenerative medicine and responsive materials, requires polymers that contain the information that is required for the materials to respond as their environment changes. Design of such sensing and responsive polymers requires the availability of molecules where the materials engineer has absolute control over the polymer composition and molecular structure. The requirement for cost-effective, precisely defined complex polymers is in conflict with existing chemical synthesis, where increases in complexity are associated with near exponential increases in time and cost of synthesis (Rodriguez-Cabello et al. 2006).

Proteins are polymers that are ideally suited as templates for functional materials. They are widely used in nature as the building blocks of structural materials for a diverse range of mechanical functions. In addition, proteins can contain large amounts of biological information. For example, type 1 collagen has more than 50 documented molecular interactions (Sweeney et al. 2008; Werkmeister and Ramshaw 2012) and tropoelastin likely has a similar number (Mithieux et al. 2013). As a protein is defined by a DNA sequence, recombinant technology can be used to generate replicas that have the same length and composition. DNA sequences can readily be modified using modern molecular biology, meaning that the increase in protein complexity is not tied to an increased time and/or cost of production.

Despite the promise of recombinant proteins as polymers for the production of advanced materials, recombinant protein polymers are currently underutilised in materials science. This is primarily due to a lack of proteins that can be (a) produced at commercial scales, (b) fabricated into useful load-bearing forms, and (c) modified to increase functionality without compromising the ability to either produce them at scale or fabricate them into material forms. Finally, there is insufficient understanding of the structure-function relationship of most proteins; knowledge that is required to unleash the true power of this approach for developing the novel and unique advanced materials of tomorrow. In this chapter we look in detail at the

challenges and advances associated with each of these issues. Finally, we outline the efforts of our laboratory towards production of a template protein polymer suitable for materials' applications.

15.2 Natural Protein Polymeric Materials: Lessons from Biology

15.2.1 Information in Structural Proteins

The concept of using structural proteins as a templates for functional information is derived from nature. In the correct environment, the amino acid sequence of natural proteins contains the information required to fold the polymer into the complex structure that is its functional form (Fig. 15.1). The primary sequences of some of these proteins are rich in bioactive information. For example, in addition to the mechanical role played by elastin and collagen in the extracellular matrix, these
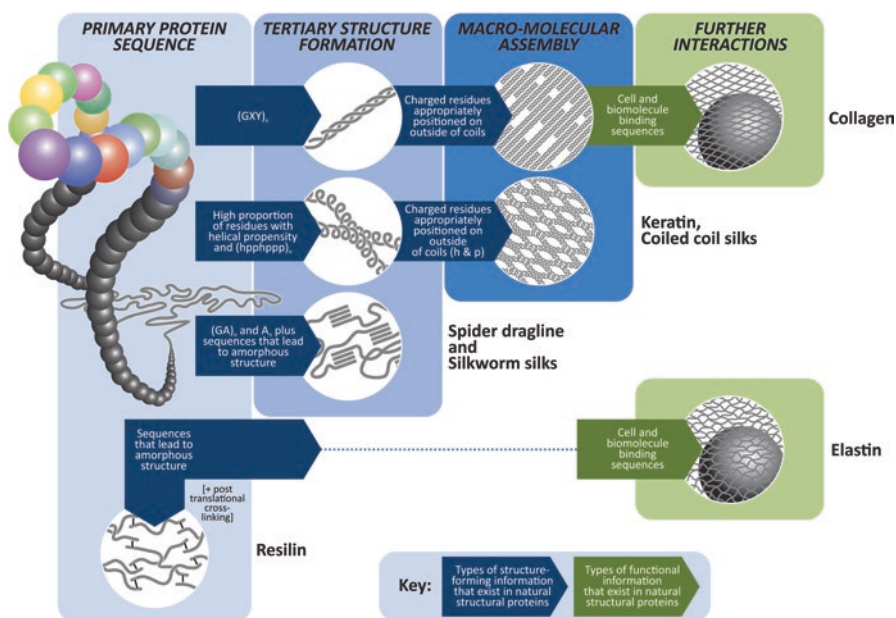
Fig. 15.1 Information within common classes of structural proteins (collagens, keratins; silks, elastin and resilin). The primary amino acid sequence contains structure-forming information to fold the proteins into their secondary structures (e.g. helices, β-sheets etc) as well as their macromolecular structure. Some classes of proteins, such as the collagens and elastin, contain addition functional information such as cell binding sites or sequences that dictate controlled protease degradation. *G*: glycine; *A*: alanine; *X*, *Y*: variable amino acid; *h*: hydrophobic amino acid; *p*: polar amino acid

proteins contain information that allows them to regulate intercellular communication, provide a scaffold for adhesion, and enable wound healing (Fig. 15.1). Based on this precedent set in nature, structural proteins have attracted considerable interest as templates for new materials. The natural structure and characteristics of some of the common structural proteins that have been used as templates for the design of novel materials are described in the sections below.

15.2.2 Elastin and Resilin

Elastin is a structural protein important in allowing the repeated extension and contraction of lungs, skin and arteries. The protein adopts an amorphous molecular structure (summarised in Muiznieks et al. 2010) held together with inter- and intramolecular cross-linking via lysine residues (Partridge et al. 1966; Davis and Anwar 1970; Siegel et al. 1970; Reiser et al. 1992; Fig. 15.1). This structure acts like an 'entropic spring', where relaxation of the forces causing extension allows the molecular structure of the material to return to an original state that has decreased order (and lower entropy) compared to the extended material. The disordered secondary structure of the elastin protein has been confirmed using a range of techniques, including circular dichroism, infrared spectroscopy and sedimentation velocity analyses (Muiznieks et al. 2010). The structural heterogeneity in the material and the high mobility of the elastin backbone has prevented collection of atomic level structural data using methods, such as high-resolution nuclear magnetic resonance (NMR) or X-ray diffraction (Fleming et al. 1980; Torchia and Piez 1973).

Like elastin, resilin is a protein that forms a rubber-like material. Resilin has excellent elasticity and fulfils a wide variety of roles in insects as an energy storage or absorption material, including allowing insects to perform repeated power jumps, absorbing shocks and vibrations, or providing a counter force for muscles. As with elastin, the structure of resilin has been shown to be highly amorphous by a range of techniques, including low and high angle x-ray diffraction and transmission electron microscopy (eg. Elliot et al. 1965). Rather than lysine cross-links, resilin is stabilised by tyrosine cross-links (Fig. 15.1; Andersen 1964).

The source of the excellent elasticity of resilin has been debated over the last 50 years, with two main theories being proposed. The commonly accepted model these days (van Eldijk et al. 2012), first proposed by Weis-Fogh (1961a, b) and backed up by studies on elastin (Hoeve and Flory 1974; Li and Daggett 2002; Pometun et al. 2004), is based on the classical statistical theory of rubber-like elasticity (Flory 1953) in which elasticity is attributed to a decrease in conformational entropy on deforming an amorphous kinetically free random polymer network. The alternative model, first proposed for elastin by Urry and co-workers (Urry 1988; Urry et al. 1995), has a molecular basis. Elastin and resilin contain several N- and C-terminal repeat sequences of VPGVG, which are predicted to fold into easily deformable β-spirals. As these repeats vary somewhat in sequence, the spiral turns are slightly irregular. This diminishes the tendency to form local crystalline regions in an easily

deformable material, thus maintaining the protein chains in a kinetically free state. Extension results in a decrease in entropy in the peptide segments, linking the β-turns due to damping of the amplitude of the rocking motions (of the librations). While this librational entropy mechanism might represent some local features of resilin and elastin, it does not provide a good and complete explanation of elasticity (Li and Daggett 2002).

15.2.3 Spider Dragline, Silkworm Cocoon and Some Insect Silks

Spider and silkworm silks have convergently evolved semi-crystalline materials that are composed of β-sheet crystallites embedded in an amorphous protein matrix. Silks with β-sheet crystallites have also evolved in other insects, including raspy crickets (Walker et al. 2012) and some sawflies (Rudall 1962). Primary sequence blocks that are high in alanine and glycine, contribute to the β-sheet regions, whilst sequence blocks high in hydrophilic residues contribute to the amorphous regions (Fig. 15.1). The proteins are large (>250 KDa), a feature that is crucial for the mechanical properties of fibres (Lin et al. 2015). The protein secondary structure content in the solid state, including conformational distributions and percentages as well as their orientation distributions along the fibre axis, has been elucidated using various spectroscopies, including mid-infrared (Lawrence et al. 2008; Sutherland et al. 2014), near-infrared (Mo et al. 2006), Raman (Lefèvre et al. 2007) and NMR (Holland et al. 2008; van Beek et al. 2000; Kameda and Tamada 2009; Kameda 2012). In the solution state, the atomic level structure of the spider silk's N- (Askarieh et al. 2010) and C- (Hagn et al. 2010) terminal domains have been determined using NMR spectroscopy and X-ray diffraction.

15.2.4 Collagen

Collagen is one of the most abundant proteins found in mammals. It is present in large quantities in skin, bone and connective tissues, providing tensile strength and elasticity. The atomic structure of the collagen molecule was initially proposed in 1954 (Ramachandran and Kartha 1954) and since then the specific covalent binding of these molecules into fibrils, which are further assembled into fibres, has been well described (Mouw et al. 2014).

Fundamental to the structure is the characteristic collagen tri-amino acid repeat, (Gly Xaa Yaa)$_n$ (Fig. 15.1). Each of these residues occupy a distinct position in the collagen helix with glycine residues buried in the core, Xaa residues being highly solvent accessible, and Yaa positions being somewhat less solvent accessible (Jones and Miller 1991). The close association of the proteins in the triple helix restricts the

core position to glycine, the smallest of the naturally occurring amino acids. In mammalian collagens, around 18 % of residues in the Xaa and Yaa position are imino acids (proline or the post-translational modification of proline to hydroxyproline), which are important for the stability of the molecule at body temperature. In bacterial collagen-like proteins (Yu et al. 2014; Persikov et al. 2005) and insect silk collagens (Sutherland et al. 2013), which do not have hydroxyproline, the imino acid distribution is different and stability is attributed to the presence of specific charged residues. In mammalian collagens, the non-glycine and imino acid residues frequently have bioactive roles that dictate binding to other collagen molecules or biological components.

15.2.5   Hair, Wool, and Some Insect Silks

The biological materials under this title all share a molecular structure that contains significant portions of α-helices that are associated into a higher order structure, which is known as a coiled coil. The coiled coil structure is one of the better-understood protein folds, with the structure readily detected by bioinformatic analysis of the primary sequence of a protein (Lupas et al. 1991). The structure is characterised by a heptad periodicity of amino acid character (not identity), with the first and fourth residues (core residues) being generally more hydrophobic (*h*) and the other residues being generally more polar (*p*) (Fig. 15.1).

The structure of hair and wool fibres consists of bundles of inner spindle-shaped cortical cells held together by a protein glue known as the cell membrane complex. This structure is surrounded by overlapping layers of thin cuticle cells in a manner similar to roofing tiles (Bradbury 1973). Prior to emerging from the follicle, keratin fibres are stabilised by a keratinisation or hardening process which involves the formation of disulfide cross-links (Marshall et al. 1991). At the lowest level of this complex hierarchical structure are microfibrils, which are comprised of a rod shaped central domain composed of four coiled coil segments with heptad repeats separated by three shorter non-helical domains (Fraser et al. 1971). The ends of the polypeptide chain consist of non-helical C- and N-terminal domains (Gillespie 1990). These coiled coils assemble further into microfibrils or intermediate filaments (IFs) which are embedded in an amorphous protein matrix. Groups of IFs form macrofibrils which run the length of the cortical cells.

The silks of aculeate insects (bees, ants and hornets) have considerable levels of coiled coil structure with the coiled coils aligned parallel with the silk fibre (Sutherland et al. 2014). The silk of different species also contains varying amounts of β-sheet structure (Sutherland et al. 2014), which is believed to cross-link the coiled coils together into a functional material. These silk proteins form hierarchically structured 'tactoids' or 'bars' in the silk gland, presumably as ordered coiled

coils, prior to silk formation (summarised in Walker et al. 2015). However, in contrast to hair and wool, the hierarchical structure in the insect silks is limited to the level of secondary and tertiary structures as the coiled coils themselves do not appear to be ordered in the silk (Walker et al. 2015).

15.3 Addition of Information into Structural Proteins

Future polymeric materials will be generated from information-rich polymers with the information within the material dictating how the material responds to changes in its environment. Although natural structural proteins are already information-rich polymers (Fig. 15.1), this information cannot be modified. A key feature of recombinant proteins is that their method of production opens the potential to insert new information into them, therefore allowing the materials scientist to design "smart" materials (DiMarco and Heilshorn 2012). By simple manipulation of the DNA sequence, new information can be introduced into the protein sequence with molecular level reproducibility. Through computational modelling, high throughput screening, and our deepening understanding of protein structure-function relationships there is an increasing knowledge of peptide motifs that can be introduced to provide new functionalities. These new motifs can be sequences of amino acids, such as proteolytic sites and known cell binding motifs, or motifs that form three-dimensional structures, such as some biomolecular binding sites.

Although in its infancy, there are already a number of examples where information has been added into structural proteins to generate new functional materials (summarised in Table 15.1). Two main approaches have been taken. In the first approach, structural domains such as elastin-like or silk-like domains are linked with a functional domain such as sequences that promote cellular adhesion (Fig. 15.2a). In this approach, the aim is for the structural domain to provide the mechanical properties while the functional domain introduces new features. To date, this is the most commonly employed tactic to design protein-based materials and allows for a mix-and-match approach for producing a range of designer chimeric proteins.

An alternative, and far less explored approach, is to insert functional motifs directly into structural domains (Fig. 15.2b). The collagen-like protein from *Streptococcus pyogenes* (ScI2) has the structural integrity of animal collagens and crucially lacks any of the 40 binding sites found in native human collagen. Therefore ScI2 represents a template for collagen protein engineering into which functional information from animal collagens can be introduced (refs in Table 15.1). More recently, Wang and co-workers developed a synthetic collagen encoding gene where natural integrin binding sites were removed to produce a collagen-mimetic polymer into which new cell binding sites could be introduced (Que et al. 2015).

Table 15.1 Examples of amino acid sequence modifications to recombinant structural proteins that have led to new functional materials

Examples of fusion of sequence motifs to structural domains

Cell binding[a]

RGD (Motif found in several extra cellular matrix proteins which stimulates cell attachment)

REDV (Cell attachment motif from fibronectin CS5 domain)

IKVAV (Laminin derived sequence that promotes cell adhesion, angiogenesis and neurite outgrowth)

Growth factor activity[b]

CNDMTPEQMATNVNCSSPERHTRSYDYMEGGDIRVRRLFCRTQWYLRIDKRGKVKGTQEMKNNYNIMEIRTVAVGIVAIKGVESEFYLAMNKE
GKLYAKKECNEDCNFKELILENHYNTYASAKWTHNGGEMFVALNQKGIPVRGKKTKKEQKTAHFLPMAIT (Human keratinocyte growth factor)

Growth factor binding[c]

CKAAKRPKAAKDKQTK (Heparin binding domain)

Proteolytic degradation[d]

GTAR↓SAS; TSHR ↓SAS; DRIR ↓SAS (Cleavage site of protease secreted from the tips of growing neurites – urokinase plasminogen activator (uPA);
uPA and tPA –sensitive domains)

GPQGJ1WGQ (Matrix metalloproteinase sensitive domain)

Transport across eukaryotic cell membranes[e]

YGRKKRRQRRR (Twin arginine translocation signalling peptide for protein export derived from HIV-1)

GRKRRQTSMTDFYHSKRRLIFSKRKP (Bac-7: antimicrobial peptide shown to promote transport of polypeptides to the nucleus)

RQIKIWFQNRRMKWKK (Penetratin – derived from the Drosophila transcription factor Antennapaedia)

AAVALLPAVLLALLA (Membrane transport sequence (MTS) – hydrophobic peptide derived from Kaposi fibroblast growth factor)

Mineralisation and metal binding[f]

SSKKSGSYSGSKGSKRRIL (R5 peptide derived from the silaffin protein of *Cylindrotheca fusiformis* induces and regulates silica precipitation)

GDNPDNTSQTGDQRDSESSEEDRLNTFSSSESQSTEFQGDSESNESLSLSEESQESAQDEDSSSQEGLQSQSASRESRSQESQSEEDSRSEENRDSD
SQDSSRSKEESNSTGSTSSSEEDNHPKNIEADNRKLIVDAYHNKPIGDQDDND CQDGY (Hydroxyapatite binding peptide derived from the
C-terminal domain of dentin matrix protein 1

MSPHPHPRHHHT; KSLSRHDHHHH (Silica binding peptides identified by phage display)

NPSSLFRYLPSD; WSWRSPTPHVVT (Silver binding peptides identified by phage display)

EQIAEFKEAFALCTKDGTGYITTKELGTCMRSLTS (Uranium recognition motif of calmodulin protein from *Paramecium tetraurelia*)

***Antimicrobial activity*[g]**

CYCRIPACIAGERRYTSGTCIYQGRLWAFCC; VCSCRLVFCRRTELRVTSGNCCLIGGVSFTYCCTRV; DTHFPICIFCCGCCHRSKCGMCCKT (Derived from human antimicrobial peptides human neutrophil defensin 2 (HNP-2), human neutrophil defensin 4 (HNP-4) and hepcidin)

***Examples of insertion of functional information into structural domains*[h]**

GFPGER GFPGEN; GERGFPGERGVE; GRPGKRGKQGQK (Introduction of intergrin; heparin and fibronectin binding sites into bacterial collagens)

GROGER, GAOGER, GLOGEN, GLKGEN, and GMOGER removed from human collagen III gene. GFOGER introduced (Integrin binding)

[a]Liu et al. (2004), Girotti et al. (2004), Heilshorn et al. (2005), Bini et al. (2006), Liu and Tirrell (2008), Charati et al. (2009), Straley and Heilshorn (2009), Nakamura et al. (2009), Wohlrab et al. (2012) and Widhe et al. (2013)

[b]Koria et al. (2011)

[c]Charati et al. (2009)

[d]Straley and Heilshorn (2009) and Charati et al. (2009)

[e]Massodi et al. (2005), (2010) and Bidwell et al. (2007)

[f]Wang et al. (2006), Huang et al. (2007), Currie et al. (2011), Mieszawska et al. (2010), Canabady-Rochelle et al. (2012), Belton et al. (2012), Krishnaji and Kaplan (2013) and Wong Po Foo et al. (2006)

[g]Gomes et al. (2011)

[h]Cosgriff-Hernandez et al. (2010), Seo et al. (2010), Peng et al. (2014a), An et al. (2014) and Que et al. (2015)
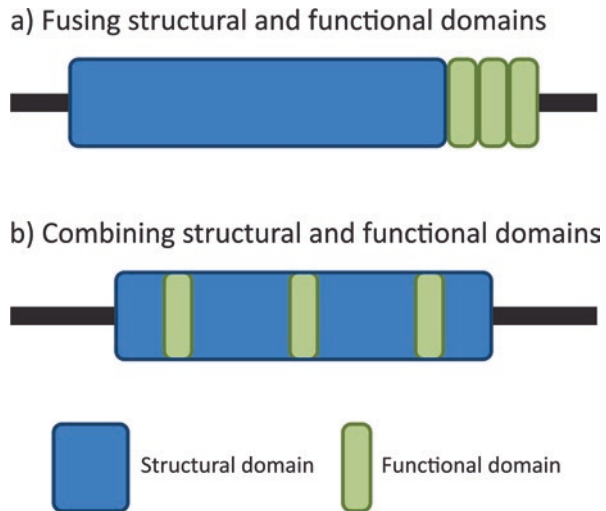
15.3.1 Examples of Designed Information-Containing Recombinant Proteins

Specific examples of primary sequence modifications to structural proteins that have conferred new properties are presented in Table 15.1. The extracellular matrix has served as one of the main inspirations for protein design with the goal of developing biomaterials for tissue engineering. Both approaches described in Fig. 15.2 have been investigated. Sequence motifs required for native functionality such as cellular adhesion, growth factor adhesion and proteolytic degradation have been introduced into collagen, or fused with elastin, and spider silk (Cell binding, Growth factor activity and binding, Proteolytic activity, Table 15.1). By introducing bioactive information, new materials with cell binding properties, or susceptibility to matrix metalloproteinase degradation have been demonstrated.

Inspired by the skeletal architectures found in nature, recombinant structural proteins have been investigated as templates for inorganic materials (Mineralisation and metal binding, Table 15.1). Here, the self-assembly properties of structural proteins are exploited with information added in the form of inorganic binding peptides to control nucleation and growth of the inorganic component. Using this approach, silica binding peptides have been fused to spider silk sequences (Wong Po Foo et al. 2006; Canabady-Rochelle et al. 2012). Through varying the length of the silk repeat sequence, from 6mer to 15mer, the morphology and physical properties of the material could be varied (Canabady-Rochelle et al. 2012). Material capable of hydroxyapatite nucleation was achieved after fragments of spider silk proteins were fused with a peptide from the dentin matrix protein 1, a protein that initiates nucleation and controls the structure and crystallographic orientation of hydroxyapatite during the mineralisation of bone and dentin. Peptides that bind other metal ions such as

uranium (Krishnaji and Kaplan 2013) and silver ions (Currie et al. 2011) have been fused to various structural proteins and the resultant material investigated for their use in bioremediation and as antimicrobial agents, respectively.

An approach to the design of structural recombinant proteins that has received a lot of attention relies on an unusual property of elastin. Elastin-like sequences (for example, $(VPAVG)_n$ and $(VAPGVG)_n$) undergo an inverse phase transition—they are soluble in aqueous environments below their critical solution temperature and abruptly aggregate above it. By manipulations of the "guest residues" of these repeat units, the phase transition can be induced under different environmental conditions (light, kosmotropic salts or pH stimuli; summarised in DiMarco and Heilshorn 2012). This property has led to these peptide sequences being incorporated in a range of protein materials to generate environmentally responsive materials. For example, a recombinant protein was designed to specifically deliver the chemotherapeutic agent doxorubicin to regions of hyperthermia (Bidwell et al. 2007). The protein contained four fused peptides, each of which were included to confer a specific functionality: cellular uptake was facilitated by the HIV-1 Tat protein, thermal targeting occurred through an elastin-like peptide, a linker site contained a cleavage site and finally a cysteine was included to allow attachment of the doxorubicin derivative. This work is an excellent example of the application-specific properties that can be achieved by introducing information into protein materials.

A further avenue that has been explored to generate materials with new properties is the fusion of naturally occurring structural repetitive sequences to produce "hybrid" materials. For example, fusion of silk and elastin-like motifs has been used to generate materials that spontaneously form nanofibrillar hydrogels under physiological conditions (Colino et al. 2014). The morphologies of silk-elastin-like proteins change as the ratio of silk to elastin motifs are varied (Xia et al. 2011). Similarly, elastin- and resilin-like sequences fused to peptides encoding collagen sequences self-assemble and form higher order fibrillar structures (Bracalello et al. 2011).

There has only been one system where functional motifs have been designed into a structural domain: rational design of bacterial collagens to resemble mammalian collagens. Mammalian collagen is one of the most 'information-rich' natural proteins known, with some molecules capable of over 50 different molecular interactions (Sweeney et al. 2008; Werkmeister and Ramshaw 2012). However, mammalian collagens contain hydroxyproline, a post-translational modification of proline that increases the thermal stability of the molecule. The requirement for hydroxyproline has severely hampered the large-scale production of these proteins in recombinant systems. In contrast, a number of bacteria genes encoding collagen-like sequences (Rasmussen et al. 2003) that fold into the collagen structure without hydroxyproline (Yu et al. 2014) have been discovered. These molecules are likely stabilised by their high content of charged residues in the X and Y positions (Mohs et al. 2007). Without the requirement for post-translational modification, these proteins are well suited for recombinant production in *E. coli*. One of these bacterial proteins, *Scl2* does not naturally contain any known cell binding sites and there has been significant effort to rationally design information into its collagen domain (Table 15.1; Examples where functional information has been inserted into structural domains).

In all cases the modified recombinant bacterial collagen folded the introduced sequences into a collagen triple helix and the resultant molecules demonstrated the introduced biological function.

Synthetic chemistry, where non-natural amino acids have been introduced into proteins, has also been used to modify the properties of structural proteins. For example, glycine has been substituted for by the non-natural amino acid aza-glycine to generate additional hydrogen bonds between the protein strands in collagen and therefore increase the stability of the molecule in solution at 55 °C (Zhang et al. 2015). Currently the commercial use of non-natural amino acids in proteins is limited by low protein yields. If issues around scalability can be resolved, then the amount of chemistry available to be incorporated into proteins will be enormous. This will be a big step to making proteins truly become the polymers of the future.

15.3.2   Constraints Limiting Rational Design of Recombinant Structural Proteins

Rational design of protein-based advanced materials requires that the material's engineer is able to predict the effect modifications of the amino acid sequence will have on the final properties. The ideal protein for materials design would be one that tolerates modifications to its primary amino acid sequence without compromising its existing properties. There are however, a number of factors that either limit where modifications can occur or that limit our ability to predict the effect of these modifications on the final material's properties.

15.3.2.1   Tolerance to Addition of Information

There are limitations to the information that can be added into recombinant structural proteins using either of the approaches described in Fig. 15.2. The amount of information that can be added is limited in the block copolymer approach, where structural domains and functional motifs are fused (Fig. 15.2a). This approach ideally aims to harvest the properties of both structure and function. It is often found that as more bioactive information is included and the relative amount of the structural domain is reduced, the mechanical properties of the material are compromised and ultimately the amount of information that is added is limited.

In contrast, incorporating information directly into the structural motif (Fig. 15.2b) seeks to fully retain the material's mechanical properties. However, no naturally occurring protein will provide a blank-slate for protein engineering as existing information dictating structure formation limits what further information can be added. The extent of this limitation depends on the class of structural protein. For example, the collagen tertiary structure requires a glycine at every third position in the protein sequence of the structural domain, as this is the only residue that can fit

in the core of the triple helical structure. The elastic properties of elastin and resilin rely on an amorphous structure and therefore if the functional sequence motifs being added to the polypeptide generate a secondary structure then the elastic properties will be compromised. In silkworm and spider silk proteins, adding sequences that disrupt hydrogen bonding between the GA and A regions will compromise the strength and stability of the material.

15.3.2.2   Absence of a Precise Amino Acid Sequence-Structure Relationship

Rational design of protein-based advanced materials requires an understanding of the three dimensional structure of the proteins in the material form. Several well-studied proteins do not have a precise relationship between their primary amino acid sequence and atomic structure in the solid state and therefore development of materials will rely on functional assessment rather than rational design. Spider and silkworm silks fall into this group. When concentrated in the silk gland, the proteins form large micelle-like mesogens (Vollrath and Knight 2001; Jin and Kaplan 2003). During silk fabrication, mechanical deformation resulting from shear forces in the silk spinning duct causes molecular extension of the proteins and brings adjacent chains into close apposition. Proximity promotes the formation of intermolecular hydrogen bonds and the formation of crystallites with β-sheet secondary structure (Jin and Kaplan 2003). There is little to distinguish individual β-sheet crystallite forming regions so crystallite formation can take place anywhere along the chain and therefore there is no precise relationship between these regions in the final material.

Proteins with amorphous structures in the solid state also do not have a predictable primary amino acid sequence-structure relationship. These proteins include the sections between the β-sheet crystallites in spider and silkworm silk, as well as the structural domains of elastin and resilin. These regions can be very extensive, for example up to 80 % of the backbone carbonyl carbon atoms of elastin have been shown to be highly mobile in polar solvents (Lyerla and Torchia 1975). Whilst not compatible with rational design, the flexible and highly dynamic backbones of these protein regions may be better suited for incorporation of information that is based on primary sequence rather than three-dimensional structure, such as proteolysis sites. Incorporation of such primary sequence motifs into these protein regions means that they are likely to be solvent exposed in solution and therefore many will be available at the surface of recombinant protein materials.

15.3.2.3   Our Understanding of Amino Acid Sequence-Structure-Function Relationship

Naturally occurring proteins are complex molecules where the relationship between the amino acid sequence, protein molecular architecture, and function has been optimised over millions of years. To date efforts to replicate the precise structure of

natural materials has been achieved to only a limited scale. Currently there is only limited understanding about these relationships and atomic level structural information for most recombinant structural proteins in solid-state materials is limited. It is undoubtedly this aspect of proteins that has most severely limited their adoption as polymers of choice for advanced materials engineering.

The primary amino acid sequences of all the structural proteins shown in Fig. 15.1 have undergone intense scrutiny from a bioinformatics point of view. There is no current method that we are aware of to use bioinformatics to predict the structure of proteins in the solid state, although some of the tertiary structures readily detected by bioinformatics (collagen and coiled coil) are expected to be retained during benign methods of materials fabrication. Circular dichroism spectroscopy is commonly used to monitor the types of secondary structure in solution and how the structure changes in response to external stimuli (i.e. pH changes, denaturation, temperature). A range of experimental techniques have been used to obtain information on the types and proportion of various secondary structures present within the solid-state materials, including Fourier transform infrared, Raman and solid state NMR spectroscopies, and x-ray diffraction. Recently micro confocal Raman spectroscopy has been demonstrated to be a powerful tool for determining conformational composition and orientation distribution within single silk fibres (Walker et al. 2013a; Church et al. 2014). None of these experimental techniques provide the atomic level structural detail required for rational design. A much clearer idea of the three dimensional structure proteins adopt when processed into solid-state materials is required before rational design alone can be used to generate new materials.

15.4   Commercial Scale Production of Protein Polymers

To develop a market for protein-based advanced materials it is necessary to produce stable, purified proteins at a commercial scale at an appropriate cost. Following the advent of cloning in the 1970s, the first commercial recombinant protein, insulin, was produced by Eli Lilly in 1982. The market for recombinant proteins, primarily therapeutic proteins for clinical applications and industrial enzymes for the treatment of food, feed, detergents, and paper-pulp, then rapidly expanded and in 2008/2009 the annual global sales of biopharmaceutical and industrial proteins exceeded 92 billion USD (Mattanovich et al. 2012). However, despite recombinant protein production being a major and expanding industry, recombinant fibrous proteins are only produced on a limited scale and generally for research applications. Often this is because structural proteins are difficult to express at high levels in current expression systems. For example, despite many years of effort, the best yields of recombinant spider silk fragments reported in the scientific literature are still less than 1 g/L (summarised in Tokareva et al. 2013).

There are many different expression platforms that have been developed to produce recombinant proteins, including a number of well-characterised bacteria, yeast and animal cell lines. A number of less conventional species are also under develop-

ment such as filamentous fungi, moulds, moss, algae, protozoa and transgenic animals (summarised in Ferrer-Miralles and Villaverde 2013). Generally, bacterial systems are the fastest with recombinant protein production typically taking 1–2 days. Yeast systems require closer to a week and mammalian cell systems can take up to a fortnight (Hacker et al. 2009). Bacterial systems are cheaper than eukaryotic systems as most bacteria can be grown on simple defined media. However, eukaryotic systems are preferred for proteins that require post-translational modifications. The type of product influences the expression platform used, with around 40 % of commercial biopharmaceutical proteins in recent years being produced by mammalian cells; 30 % from the bacterium, *Escherichia coli*, and 20 % from yeasts (Martínez et al. 2012). In contrast, commercial industrial enzymes, where volume is generally critical, are predominantly produced by *E. coli or* yeast if they are recombinant proteins or *Bacillus* species or filamentous fungi if they are native proteins. To date, the yields observed in recombinant systems rarely reach the yields obtained from native protein production. Below we describe commonly used expression systems and the advantages and disadvantages associated with each for bulk protein production.

15.4.1 *Protein Production Platforms for High Level Expression*

By far the most commonly used high-level recombinant expression system is fermentation in the bacterium *E. coli*. This system has been used to successfully produce a wide range of non-modified proteins less than 60 kDa in size and is responsible for around 50 % of all recombinant protein products approved by the US FDA (Walsh 2010). The bacterium is easy to transform, has unparalleled growth kinetics with a doubling time of 20 minutes, can grow to high density (theoretically up to 200 g dry weight/L) and is capable of ultra-high yields of recombinant protein with greater than 50 % of total cell protein being possible (Rosano and Ceccarelli 2014).

The non-pathogenic bacteria, *Pseudomonas fluorescens*, can also achieve high recombinant protein titres (Chen 2012). As with *E. coli*, this species can grow to high cell densities (>100 g/L) and proteins can be produced at greater than 50 % total cell protein content (Chen 2012). Yields of 20 g/L of soluble cowpea chlorotic mottle virus-like particles, used as a vaccine conjugate, have been reported (Zhao et al. 1995; Phelps et al. 2007).

Currently, these bacterial systems have two disadvantages as large-scale expression platforms: endotoxin contamination and complex downstream purification resulting from contamination of host cell proteins. Endotoxins, or lipopolysaccharides (LPS), are major components of the outer membrane of all Gram-negative bacteria that are continuously released during cell growth, division and death. The presence of endotoxins in host organisms elicits toxic responses and hence they

need to be removed before materials can be used for *in vivo* applications. Removal of endotoxins is expensive and no existing method completely removes them from recombinant protein expressions. Endotoxin levels can be significantly reduced if proteins are expressed into *E. coli* inclusion bodies, which can readily be purified from soluble host cell proteins and contaminating endotoxins (Magalhäes et al. 2007). Alternatively, endotoxin free *E. coli* strains that contain the non-toxic endotoxin precursor, lipid Iv$_A$, as the only LPS related molecule in the outer membrane have been developed to alleviate this issue (Mamat et al. 2015).

Bacillus are commercially important bacterial species that produce enzymes such as amylases and alkaline proteases for detergents as well as for the food and agriculture industries. There are three commonly used strains: *B. subtilis*, *B. megaterium* and *B. brevis*. All three are non-pathogenic, do not produce endotoxins, have simple fermentation processes, and are capable of secreting ultra-high levels of protein (Dong and Zhang 2014). For example, reported yields include up to 20 g/L protein from *B. subtilis* (Schallmey et al. 2004) and up to 30 g/L protein from *B. brevis* (Takagi et al. 1989). *B. brevis* and *B. megaterium* produce low levels of extracellular protease in comparison with *B. subtilis*, and *B. megaterium* does not secrete alkaline proteases (Schumann 2007). Despite their potential as protein producers, as a result of a combination of poor membrane trafficking and/or translocation, poor protein folding, protease activity and/or inefficient cell wall passage (reviewed in Schumann 2007) yields of recombinant proteins are usually low. Furthermore, strain development is time consuming and the selection process is laborious (Dong and Zhang 2014). It is anticipated that much effort will be required to modify these systems before yields of heterologous recombinant proteins will reach those of homologous proteins (Schumann 2007; Dong and Zhang 2014).

The yeast *Saccharomyces cerevisiae* is a system widely used for biopharmaceutical proteins. It has protein folding and post-translational modification pathways well suited to making recombinant eukaryotic proteins and secretes proteins into the medium, thus facilitating downstream processing. It has been used to produce commercially important proteins such as insulin, human serum albumin and vaccines. However, protein yields are relatively low compared to other systems: one of the best has been 1.6 g/L of a cutinase (Calado et al. 2004). High protein yields have been reported in another yeast, *Picia pastoris* (for example 12 g/L; Clare et al. 1991). However, these levels were achieved using cytoplasmic expression and recombinant protein recovery requires cell lysis, a high-energy process for yeast. For both *Picia* and *Saccharomyces* the secretion pathways are complex and need to be better understood before recombinant protein yields can be expected to achieve those observed in the most prolific bacterial systems.

Recently there have been dramatic advances in the yields obtained by mammalian cell lines with Chinese hamster ovary cells, the preferred mammalian cell line, reaching g/L yields (Lai et al. 2013). However, the cell line development process is expensive and time consuming and therefore advances in this area are required before this platform is suitable for general protein polymer production. Low protein yields and/or complex downstream processing, limits production of proteins in plants and cell free systems and hence, without major technological advances, it is

unlikely that these systems will be cost effective for commercial scale production of protein polymers for advanced materials production.

15.4.2 Downstream Processing Considerations

Of course, high protein yield is not the only factor that is required from the expression platform. Proteins that are to be used for advanced materials need to be purified to the extent that contaminants do not interfere with protein folding and assembly into materials. Downstream processing is simplified, and therefore more cost effective, if the recombinant proteins are secreted into the medium: it is estimated that downstream processing accounts for 10–30 % of the cost for extracellular produced proteins whereas it accounts for 70–95 % of the total cost of protein production from cytoplasmic expression (Mattanovich et al. 2012). Therefore, systems that involve lower expression of secreted proteins can be cost competitive to systems with higher cytoplasmic expression and associated complex downstream purification pathways.

Secretion into the medium is not the only route to simplified downstream purification. The downstream process can also be made more efficient by identification of specific protein characteristics that allow easy purification. For example, *E. coli* has been used to produce bacterial collagen proteins at yields of up to 19.3 g/L (Peng et al. 2012). The recombinant collagens have a high isoelectric point, which allows easy purification from the host cell proteins that are readily precipitated by acid treatment (Peng et al. 2014b). Alternatively, in the absence of features that differentiate the recombinant protein from host cell proteins, recombinant proteins can be expressed into the inclusion bodies of *E. coli* to give a higher recombinant protein: host cell protein ratio. This approach requires the proteins to be amenable to unfolding from the inclusion bodies and refolding to their native form.

The level of existing commercial protein production means that the infrastructure required for scale up from the laboratory to commercial production already exists, albeit these levels are significantly lower than those available for many synthetic polymeric materials. What is currently limiting is the availability of structural protein polymers that can be expressed at high levels with properties that allow them to be purified at low cost. Progress can be made by advances in recombinant systems that allow ultra-high level expression and low cost purification of existing proteins or by the discovery and optimisation of new proteins so that they can be produced at ultra-high levels and easily purified.

15.5 Formation of Proteins into Materials

Production of advanced materials from proteins requires that the proteins can be fabricated into load-bearing material forms. In nature, proteins are fashioned into the shape and form required to fulfil the desired function; thus spiders and

silkworms form fibres to produce webs and cocoons. In some cases, these same proteins can be formed into films to strengthen the inter-connections between web filaments (Church et al. 2014). In many cases, such as in hair and wool fibres, the organism self-assembles the proteins into a complex hierarchical structure that contributes greatly to the final properties of the material.

For recombinant proteins it is possible, but unusual, to use a surrogate host to directly produce a useful product. Teulé et al. (2012), for example, used silkworms as surrogate hosts for spider silk production and generated composite silkworm/spider silk fibres with slightly improved mechanical properties compared to unmodified silkworm fibres. More common is the use of some post-production technology such as casting, spraying or spinning (Fig. 15.3), coupled with a stabilisation strategy, to impart structure and cross-links and thereby turn the protein into a useful material. Reproducing nature's often complex hierarchical structure has so far proved impossible; nevertheless sufficient structure can be developed to generate the required properties for advanced materials.

15.5.1 *Processing*

Processing generally starts by dissolving the protein in a suitable solvent. Most of the work describing processing of structural proteins into material forms has centred on proteins derived from natural silks. These silks require a strong hydrogen bond-breaking solvent in order to dissolve the proteins. Such solvents can be classified into four types according to their composition: inorganic salt systems, fluorinated organic solvents, concentrated acids, or organic salt systems (Fu et al. 2009). Generally recombinant proteins are more readily soluble in benign aqueous solutions as, after expression and before processing into materials, they do not have the extensive inter-protein hydrogen bonding or covalent networks present in natural materials. There are several excellent reviews describing advances in the development of protein based biomaterials which include descriptions of current methods to process protein solutions into films, fibres, hydrogels, spheres, capsules and sponges or foams (Hardy et al. 2008; Hardy and Scheibel 2009, 2010; Vepari and Kaplan 2007; Fu et al. 2009; Humenik et al. 2011; van Vlierberghe et al. 2011; Tjin et al. 2014). Rockwood et al. (2011) give details of the protocols required to fabricate hydrogels, tubes, sponges, composites, fibres, microspheres and thin films from *B. mori* silk.

15.5.1.1   Wet Spinning

The process of extruding a polymer solution or dope through a spinneret to form a continuous filament is called wet spinning. The fibres are generally extruded directly into a bath that causes the protein to come out of solution as a fibre which solidifies as it emerges from the bath. For proteins the dope is often an aqueous solution and
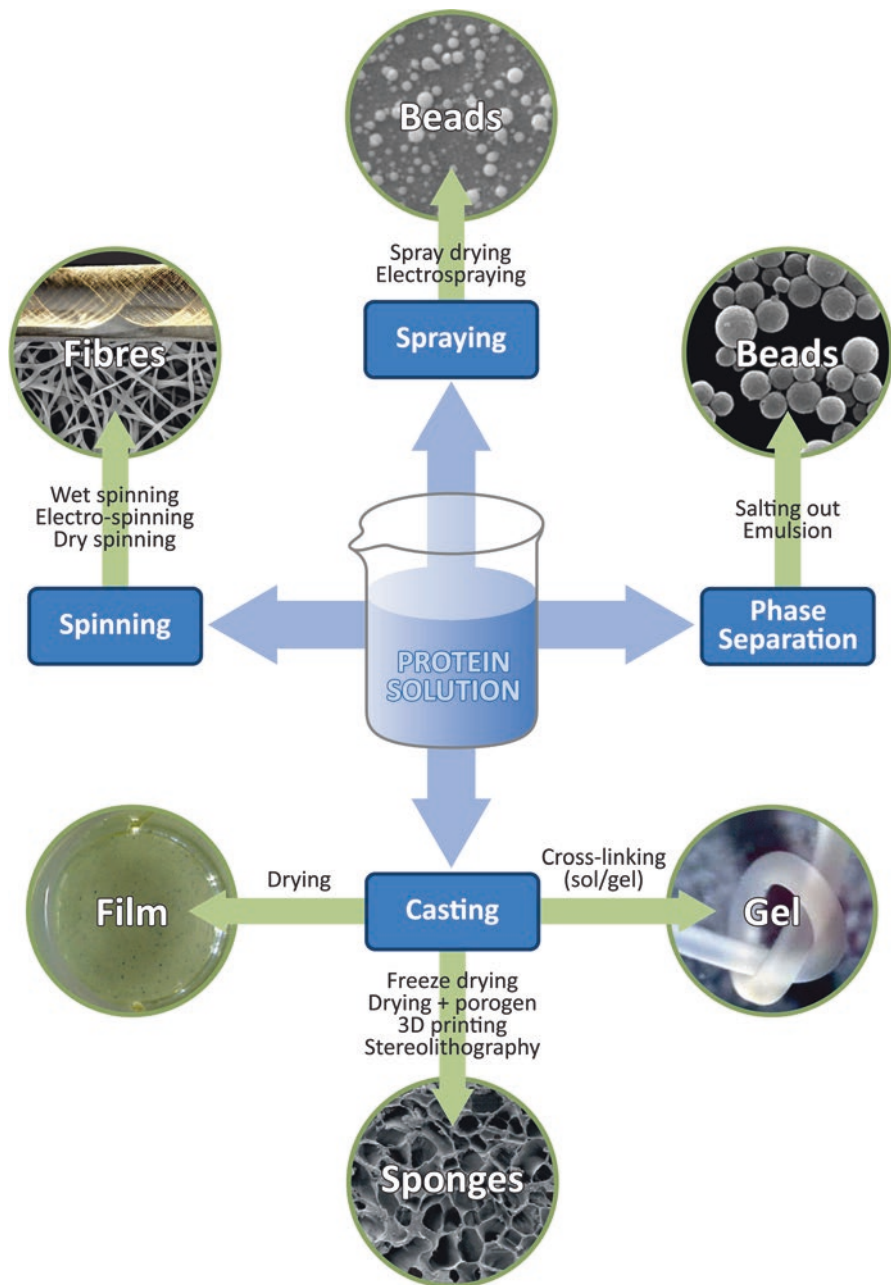
Fig. 15.3 Different processing methods for converting a protein solution into microspheres, beads, films, hydrogels, sponges and fibres (Images courtesy of Elvin et al. 2005; Jiankang et al. 2007; Slotta et al. 2008; Hardy et al. 2008; Gholami et al. 2011; Fan et al. 2012; and Poole et al. 2013)
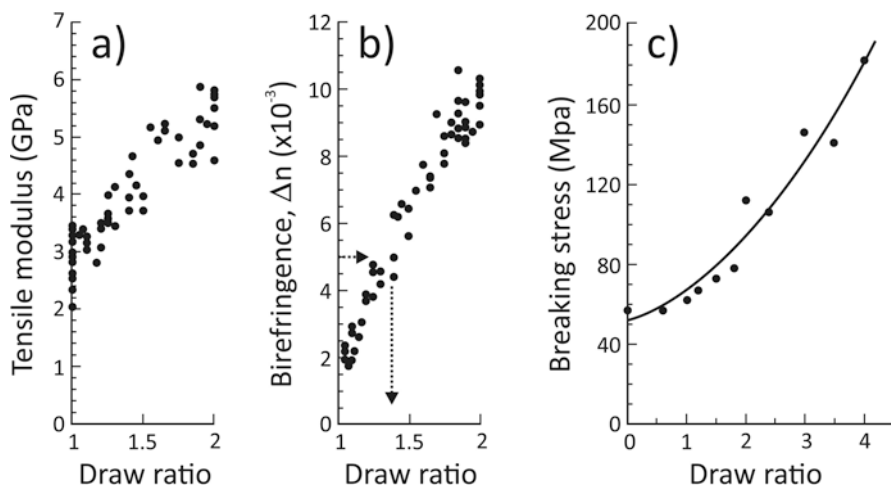
Fig. 15.4 Effect of draw ratio on the tensile modulus (**a**) and birefringence (**b**) of reconstituted dried hornet silk gel film (Reprinted with permission from Kameda et al. 2010. Copyright 2010 American Chemical Society) and on the breaking stress (**c**) of recombinant honeybee silk fibres (our data, unpublished)

the coagulation bath an alcohol-water solution. Once solidified, the fibres can be drawn to induce protein chain orientation in the fibre direction, in order to improve the mechanical properties (Kameda et al. 2010; Sutherland et al. 2011). Figure 15.4 shows the increase in physical properties of reconstituted dried hornet silk film (Kameda et al. 2010) and honeybee silk fibres as the extent of drawing increases. There are many examples where wet spinning has been used to produce fibres, including from recombinant honey bee silk (Poole et al. 2013), recombinant spider silk (Xia et al. 2010; Humenik et al. 2011; Gnesa et al. 2012; Tokareva et al. 2014) as well as reconstituted silkworm silk fibroin (Marsano et al. 2005; Pérez-Rigueiro et al. 2009; Zhou et al. 2009; Jin and Kaplan 2003; Yan et al. 2010; Zhou et al. 2014) and spidroin (Seidel et al. 1998; Seidel et al. 2000; Shao et al. 2003). The effects of spinning from different solutions are discussed in detail by Fu et al. (2009). In some cases fibres have been air spun by drawing straight from a concentrated solution of the protein (Shao et al. 2003; Xie et al. 2006; Weisman et al. 2010).

15.5.1.2 Electrospinning

Electrospinning is a wet spinning process that uses an electrical charge to draw very fine fibres from a polymer solution placed in a strong electric field. Typically the solution is introduced by pumping through a needle or spinneret, but recently needleless electrospinning setups have been reported. They are designed to increase the nanofiber production rate. These new devices include the use of rotating rollers (Dosunmu et al. 2006; Varabhas et al. 2008; Yener and Jirsak 2012) and spirals

(Wang et al. 2012) as the fibre generator. The fibres dry in flight and are normally collected as a non-woven mat on a grounded collector electrode. Variations on this theme allow fibres to be collected on a support fabric (Heikkilä et al. 2007), coated onto a yarn (Dabirian et al. 2012), or directly spun into a yarn (Ali et al. 2012).

The spinning characteristics and resultant fibres are dependent on a range of parameters such as the viscosity and conductivity of the polymer solution, applied voltage, collector type, spinneret-collector distance, flow rate, solvent system (evaporation rate) and temperature. The advantage of electrospinning is that a useful sample can be made with a very small quantity of material. The extremely fine diameter of the fibres results in a large surface area, and work on single (Arinstein et al. 2007) as well as aligned bundles (Hwang and Yu 2009) of nanofibres showed they are intrinsically much stronger than their microfibre counterparts. Fu et al. (2009) give a good overview of the structure and properties of different silk fibres electrospun under a range of conditions, whereas Li and Wang (2013) cover the effect of varying the solution (viscosity, conductivity, etc.) and processing (applied voltage, flow rate, etc.) parameters.

15.5.1.3 Casting of Films, Gels and Sponges

The easiest way to turn a protein solution into a useful material is to pour it onto a smooth surface and allow the solvent to evaporate, thus forming a film (Kameda et al. 2005; Hu et al. 2006; Lawrence et al. 2002, 2008; Hu et al. 2011; Rajkhowa et al. 2011; Humenik et al. 2011; Qin et al. 2012). To form a more uniform film the solution can be spin-coated where the substrate is rotated at high speed in order to spread the coating material by centrifugal force. Whilst many protein films need to be treated to impart stabilisation to water (Hu et al. 2006, 2011; Lawrence et al. 2002, 2008), some self-assemble as the concentration of polymer in solution increases during drying, forming structures that impart water stability to the film (Lu et al. 2010). The use of different solvents can also affect the final film structure. Recombinant peptides of spider silk cast from hexafluoroisopropanol solution form α-helical and random coil molecular structures and the film readily dissolves in water whereas when cast from formic acid solution the β-sheet rich films formed are stable in water (Humenik et al. 2011).

A protein solution can also be poured into a mould and cross-linked in solution to form a hydrogel. Hydrogels are networks of hydrophilic chains where water is the dispersing medium. This has been done to good effect with recombinant resilin, which is readily cross-linked using blue light (Elvin et al. 2005). Wang et al. (2008) subjected silk fibroin solution to ultrasonication to accelerate the gelation of the solution by initiating the formation of β-sheets. Many biopolymer solutions gel spontaneously when cooled below their upper critical solution temperature (van Vlierberghe et al. 2011), the gelation being driven by the formation of structures such as double helices (for polysaccharides) or triple helices (for gelatin).

Several methods exist for converting protein solutions into sponges or foams, the most common being lyophilisation (Nazarov et al. 2004; Huson et al. 2012) or the

use of porogens, which are subsequently either leached out (salt leaching) or subli-mated (gas foaming) (Nazarov et al. 2004). Lyophilisation, or freeze-drying of the protein solution to form a sponge, is followed by stabilisation (see section on cross-linking). In some cases freezing itself induces a structural stabilisation as a result of an increase in the concentration of polymer in solution when a portion of the pure solvent crystallises (van Vlierberghe et al. 2011). These materials are often referred to as cryogels. In non-aqueous solvents, porogens such as salt or sugar can be used. After evaporation of the solvent, the water-soluble porogen can be leached out to form the sponge. When a porogen such as NH_4HCO_3 is used, the pores are formed by the sublimation of the porogen particles upon heating. Since the pore sizes in the sponges reflect the size of the porogen used in the process (Nazarov et al. 2004) the sponge architecture can be carefully controlled. More novel methods which give precise control of the scaffold architecture have also been used, for example 3D printing of silk fibroin (Ghosh et al. 2008; Jose et al. 2015) and collagen (Kim et al. 2009) solution were used to fabricate tissue scaffolds, whilst Jiankang et al. (2007) employed stereo-lithography to prepare moulds with intricate detail which were subsequently used to create chitosan/gelatin scaffolds with predefined internal microstructures.

15.5.1.4   Spraying and Phase Separation

Microspheres or beads are generally produced using either spraying or phase sepa-ration techniques. Solutions can be spray dried (Hino et al. 2003; Genc et al. 2009; Kim et al. 2015a) or electrosprayed (Gholami et al. 2011; Kim et al. 2015b) to form particles that range in diameter from nanometres to hundreds of micrometres with varied shapes, sizes, size distributions and surface topographies. Spray drying has the advantages of negligible degradation of heat-sensitive molecules, rapid produc-tion and controlled particle characteristics with particle size generally in the 1–5 μm range. The disadvantages are the difficulty of obtaining a narrow size distribution and the tendency of particles to agglomerate (Genç et al. 2009). Electrospraying uses the same equipment as electrospinning, with beads rather than fibres forming if the polymer concentration, and hence viscosity, is below a certain threshold (Gholami et al. 2011; Li and Wang 2013). Particles can be < 1 μm in diameter and generally have a more uniform spherical shape and size when compared with other techniques reported in the literature (Gholami et al. 2011).

Wang et al. (2010) used the spontaneous phase separation between polyvinyl alcohol and silk fibroin to form emulsions, which can be dried to films. Wang et al. (2007) have also used an unsaturated fatty acid lipid to stabilise an emulsion of silk fibroin. This latter procedure lends itself to the encapsulation of drugs in the micro-spheres. Freeze-thaw cycles and sonication of the emulsion can be used to form small vesicles with homogeneous size distributions. For recombinant spidroin, phase separation has been induced by the addition of salt (Slotta et al. 2008). Rapid mixing of equal volumes of protein solution (1 mg/mL) and potassium phosphate buffer (1 M, pH 8.0) rapidly generates microspheres which after 120 seconds

develop sufficient β-sheet structure to be stable in water. Sphere size has been shown to be a function of protein concentration and mixing intensity (Lammel et al. 2008).

15.5.2   Cross-Linking

In nature, hydrogen bonding is frequently used to cross-link structural proteins to stabilise materials. The positions of these bonds have been optimised over many years of evolution to generate materials with remarkable properties (Vollrath and Porter 2006). Where proteins are not closely packed and therefore extensive hydrogen bond networks are not possible, i.e. amorphous and hydrated proteins such as resilin and elastin, covalent cross-linking is added through post-translational modifications of the protein sequences.

Whilst aqueous solutions of recombinant proteins can readily be converted into solid materials such as films, fibres and sponges by simply allowing the water to evaporate, or inducing coagulation by addition of a non-solvent, these materials generally lack the structure of natural proteins with the consequence that they invariably have poor mechanical properties and are readily dissolved in water. The absence of adequate cross-linking has often been the stumbling block in the production of useful recombinant protein materials. Fortunately, because of their physical and chemical structure, proteins can be readily cross-linked in a number of ways.

15.5.2.1   Reaction of Chemical Groups

The formation of covalent interchain cross-links requires amino acid residues with reactive chemical groups. Cysteine residues are particularly useful as they can spontaneously combine to form cystine. Other reactive residues include tyrosine, glutamic and aspartic acid, arginine, lysine, and serine (Ziegler 1977; Elvin et al. 2005). Unlike cysteine, these other residues generally require enzymes or the application of heat, light, or chemicals to generate cross-links.

Materials from reconstituted collagen have been stabilised by heating under vacuum to temperatures of 100–120 °C for several days (Yannas and Tobolsky 1967), during which isopeptide (amide) cross-links are formed as a result of condensation reactions between amine and carboxylic acid side groups. A small amount of lysinoalanine is also formed by the dehydration of serine (or desulfation of cysteine) followed by reaction with lysine (Gorham et al. 1992). Recombinant honeybee silk has also been shown to form covalent cross-links when subjected to dry heat (Huson et al. 2012). Loss of lysine and marginal loss of serine and threonine in these dry heated sponges suggest the formation of lysinoalanine and/or methyllysinoalanine cross-links, respectively. Asparagine, aspartic acid and glutamic acid levels were also diminished in the heat-treated material, consistent with the formation of isopeptide and ester cross-links, the presence of the former being verified by liquid chromatography tandem mass spectrometry.

Subjecting recombinant resilin to blue light in the presence of a metal catalyst results in cross-linking via tyrosine residues (Elvin et al. 2005). Recombinant resilin can also be cross-linked by *Arthromyces ramosus* peroxidase, an enzyme known to catalyse dityrosine formation (Malencik and Anderson 1996). Other enzymes such as transglutaminase are capable of forming isopeptide cross-links between lysine and glutamine residues (Gerrard et al. 1998) and are often used in commercial wool finishing. Glutaraldehyde forms covalent links between lysine and tyrosine residues and is often used in food applications (Gerrard et al. 2003a, b). Aldehydes can polymerise by condensation and thus extended chain cross-links can be formed.

15.5.2.2   Self-Assembly

An alternative to forming a covalent cross-link is the formation of a structural entity, which can act as a physical cross-link. The stabilisation of proteins by self-assembly of molecules to form ordered structures is prevalent in nature. In wool fibres, α-helical structures combine to form coiled coil dimers, which in turn are the building blocks of the intermediate filaments that give wool fibres their strength. The triple helix of collagen and the β-sheet structure of the silk of spiders and silkworms are classic examples of proteins being stabilised by structure.

This method of stabilisation also plays a role in recombinant or regenerated proteins. Silkworm fibroin material (Tsukada et al. 1994; Nazarov et al. 2004; Chen et al. 2007; Lu et al. 2010; Hu et al. 2011) as well as recombinant honeybee silk (Huson et al. 2012) are known to be stabilised by immersion in methanol/water solutions. Stabilisation occurs as a result of a partial molecular transition from random coil or coiled coil to β-sheet structure, with the β-sheet structure serving to tie adjacent molecules together to form a 3D network. Water acts as a plasticiser, lowering the glass transition temperature of the protein and thus increasing the mobility of the protein, whilst the alcohol induces formation of β-sheets by increasing the hydrophobicity of the solvent and thereby weakening internal protein-protein hydrophobic interactions (Huson et al. 2012). At the same time it decreases the availability of water for hydrogen bonding thereby driving protein-protein hydrogen bonding (Hwang et al. 2011). Levels of up to 60 % β-sheet crystallinity have been reached in silkworm fibroin using methanol/water, water annealing and steam sterilisation; either individually or in combination (Hu et al. 2006, 2011; Lawrence et al. 2008; Lu et al. 2010; Park et al. 2010).

15.6   Silk Proteins Studied in Our Laboratory

On a quest to identify proteins that fulfil the necessary criteria required for templates for advanced materials design, our laboratory investigated a wide number of naturally occurring structural materials (Fig. 15.5; see refs within Sutherland et al. 2010). Most of our effort now focuses on the silk proteins of the aculeate insects

New functional materials

Fig. 15.5 The process used in our laboratory to generate new functional materials. Initially, proteins suitable for high level protein expression and with amino acid sequence that are tolerant of substitutions are identified from nature. The proteins are produced recombinantly and assessed for their ability to be fabricated into material forms. Proteins that fulfil these criteria are then rationally designed to add function

(bees, ants, hornets; Sutherland et al. 2014). These proteins fold into α-helices that self-assemble into coiled coil molecular structures, which are markedly different from the β-sheet structures adopted by the silk proteins of silkworms and spiders (Rudall 1962; Sutherland et al. 2007).

The proteins are 300–600 amino acids in length (Campbell et al. 2014) with low levels of repetition in the primary amino acid sequence and hence are well suited to existing recombinant production systems. The honeybee proteins have been produced at a high level in *E. coli* laboratory systems (>2 g/L; Weisman et al. 2010) and at much higher levels (10 g/L) in commercial facilities. The proteins are expressed into the inclusion bodies enabling them to be readily purified from the majority of the host cell proteins and endotoxins (Weisman et al. 2010). Inclusion body expression is feasible because the proteins readily refold into their native coiled coil structure after being solubilised from the inclusion bodies using detergent (Walker et al. 2013b). The protein's readiness to refold possibly reflects their native function as silk proteins which are required to remain soluble at high concentration in the insect's silk gland prior to silk fabrication.

Solutions of the recombinant proteins can be formed into a variety of solid-state materials such as sponges, films and fibres, with the physical properties of the final material being determined by the fabrication process (Huson et al. 2012; Poole et al.

2013; Rapson et al. 2014). These materials can be stabilised using either a dry heat treatment, which induces covalent cross-links (Huson et al. 2012), or by aqueous methanol, which results in formation of protein-protein H-bond cross-links in the form of β-sheets (Huson et al. 2012). In these stabilised materials the coiled coil tertiary structure is preserved. Within the artificial materials the coiled coils are anisotropic, with some alignment occurring upon drawing of fibres (Poole et al. 2013). The coiled coils found in nature contain the information required to assemble into higher order structures: the coiled coils are aligned along the direction of the fibre in native silks and optical and electron microscopy studies have shown that the proteins in a number of species are highly ordered within the silk gland prior to extrusion (Flower and Kenchington 1967; Silva-Zacarin et al. 2003; Zara and Caetano 2002; Gomes et al. 2004). The conditions that promote this macromolecular assembly have not yet been elucidated and as yet assembly of the coiled coils into higher order structures has not been achieved in artificial systems.

The protein sequences from a number of aculeate silk proteins have been characterised (Sutherland et al. 2007; Shi et al. 2008; Campbell et al. 2014; Sezutsu et al. 2007), and more are available from genomic databases. Within the approximately 210 amino acid long coiled coil region, which comprises the majority of the protein sequence, there is substantial levels of sequence variation. For example, the sequences from ant/hornet species only share around 40 % sequence identity to the homologous proteins from bees (Campbell et al. 2014). The natural primary sequence variation found between homologues indicates that the sequence can be extensively modified and yet will retain the ability to self-assemble into the native coiled coil molecular structure, which can then be fabricated into silk materials. The extensive sequence variation between these proteins is not unexpected as the motif that dictates the coiled coil structure is consecutive heptad repeats of $(hpphppp)_n$ where 'h' are generally hydrophobic residues and 'p' are generally polar residues. This motif is reliant on amino acid character rather than identity and can accommodate significant variations in this general pattern. For example, multiple charged residues are found in the 'h' positions; most likely these do not disrupt the coiled coil because of the extensive length of the coiled coil domain. As long as the main properties that provide stability to the coiled coil, such as helical propensity, hydrophobicity of the core residues, tightness of core packing and favourable polar and ionic interactions (Lupas and Gruber 2005) are retained, then information can be included through the mutation of amino acid residues without disrupting the overall protein structure.

Our ability to produce these silk proteins at high levels in recombinant systems, to fabricate them into various material forms and to modify the primary protein sequence makes them ideal templates for the design of advanced materials. Attempts to obtain an atomic level model of these proteins have proven frustrating with extensive efforts at crystallisation being fruitless: crystals have been obtained but these diffract in one dimension, suggesting formation of fibrils rather than crystals. However, the coiled coil structure is one of the best-understood protein tertiary structures, and can be predicted with good accuracy from bioinformatics analysis. Based on bioinformatic predictions of atomic structure, we have recently engi-

neered a heme-binding site into materials generated from these proteins to produce sensitive and highly stable nitric oxide sensors or recoverable catalysts with peroxidise activity (Rapson et al. 2015).

15.7   Conclusions

Naturally occurring structural proteins are 'information-rich' molecules. They contain information within their amino acid sequence that drives the atomic level structure of the molecule and, in some cases, additional information that dictates how it interacts with its environment. Based on the precedent set by nature, these structural proteins provide a rich landscape for the rational design of advanced functional materials with a wide range of mechanical properties. Advances in molecular biology make it possible to readily modify the primary sequence of recombinant proteins once their corresponding DNA sequence is known and the infrastructure is in place for producing commercial levels of recombinant proteins, albeit at low scale. However, there are a number of key challenges that need to be overcome before we can repurpose these proteins for novel functions.

The complexity of these proteins presents challenges with respect to our ability to precisely design the proteins to confer materials with new functional properties. To date, 'designer proteins' have primarily involved fusion of structural domains with functional motifs. Whilst this approach has successfully generated new materials, the amount of new information that can be included is limited before the structural/mechanical integrity of the material is affected. Incorporation of information directly into the structural motifs is required for rational design of complex multifunctional materials. This direct approach has been demonstrated with insertion of bioactive motifs into bacterial collagen proteins.

A major limiting factor for rational design of new materials based on proteins is our understanding about how the amino acid sequence is related to the structure and function of the materials that we can fabricate. There is limited or no atomic level structural information for recombinant structural proteins in the solid-state. Also limiting the scope of rational design is the existing information within naturally occurring structural proteins. Each molecular structure already contains the information within its primary sequence that leads to the protein's native structure. This information must be retained if the benefit conferred by this natural structure is to be exploited and hence each protein structure has different tolerances to the insertion of new information. No naturally occurring protein will provide a true blank-slate for rational design and each protein class offers different solutions for engineering.

There are a number of recombinant protein systems that can produce structural proteins at commercially viable levels although there is a need for more proteins that are compatible with these systems. Alternatively, improvements on existing expression systems will widen the number of structural proteins that can be produced at these levels. Fabrication of advanced materials generated from rational

design will likely be reliant on the designed proteins being mono-disperse, and hence predictable in behaviour during fabrication. Therefore, there is a need for improvement in recombinant protein purification methodologies before the level of purity required for rational materials design is achieved.

Recombinant proteins offer the control of length and composition required for advanced materials engineering. We are only at the beginning of the journey in the development of designer protein-based materials.

References

Ali U, Zhou Y, Wang X, Lin T (2012) Direct electrospinning of highly twisted, continuous nanofiber yarns. J Text I 103:80–88

An B, Abbonante V, Yigit S, Balduini A, Kaplan DL, Brodsky B (2014) Definition of the native and denatured type II collagen binding site for fibronectin using a recombinant collagen system. J Biol Chem 289:4941–4951

Andersen SO (1964) The crosslinks in resilin identified as dityrosine and trityrosine. Biochim Biophys Acta 93:213–215

Arinstein A, Burman M, Gendelman O, Zussman E (2007) Effect of supramolecular structure on polymer nanofibre elasticity. Nat Nanotechnol 2:59–62

Askarieh G, Hedhammar M, Nordling K, Saenz A, Casals C, Rising A, Johansson J, Knight SD (2010) Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. Nature 465:236–238

Belton DJ, Mieszawska AJ, Currie HA, Kaplan DL, Perry CC (2012) Silk–silica composites from genetically engineered chimeric proteins: materials properties correlate with silica condensation rate and colloidal stability of the proteins in aqueous solution. Langmuir 28:4373–4381

Bidwell GL, Fokt I, Priebe W, Raucher D (2007) Development of elastin-like polypeptide for thermally targeted delivery of doxorubicin. Biochem Pharmacol 73:620–631

Bini E, Foo CWP, Huang J, Karageorgiou V, Kitchel B, Kaplan DL (2006) RGD-functionalized bioengineered spider dragline silk biomaterial. Biomacromolecules 7:3139–3145

Bracalello A, Santopietro V, Vassalli M, Marletta G, Del Gaudio R, Bochicchio B, Pepe A (2011) Design and production of a chimeric resilin-, elastin-, and collagen-like engineered polypeptide. Biomacromolecules 12:2957–2965

Bradbury JH (1973) The structure and chemistry of keratin fibers. Adv Prot Chem 27:111–211

Calado CRC, Ferreira BS, da Fonseca MMR, Cabral JMS, Fonseca LP (2004) Integration of the production and the purification processes of cutinase secreted by a recombinant *Saccharomyces cerevisiae* SU50 strain. J Biotechnol 109:147–158

Campbell PM, Trueman HE, Zhang Q, Kojima K, Kameda T, Sutherland TD (2014) Cross-linking in the silks of bees, ants and hornets. Insect Biochem Mol Biol 48:40–50

Canabady-Rochelle LLS, Belton DJ, Deschaume O, Currie HA, Kaplan DL, Perry CC (2012) Bioinspired silicification of silica-binding peptide-silk protein chimeras: comparison of chemically and genetically produced proteins. Biomacromolecules 13:683–690

Charati MB, Ifkovits JL, Burdick JA, Linhardt JG, Kiick KL (2009) Hydrophilic elastomeric biomaterials based on resilin-like polypeptides. Soft Matter 5:3412–3416

Chen R (2012) Bacterial expression systems for recombinant protein production: *E coli* and beyond. Biotechnol Adv 30:1102–1107

Chen X, Shao Z, Knight DP, Vollrath F (2007) Conformation transition kinetics of *Bombyx mori* silk protein. Proteins 68:223–231

Church JS, Woodhead AL, Walker AA, Sutherland TD (2014) A comparison of convergently evolved insect silks that share β-sheet molecular structure. Biopolymers 101:630–639

Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K, Romanos MA (1991) High level expression of tetanus toxin fragment C in *Pichia pastris* strains containing multiple tandem repeat integrations of the gene. Biotechnology 9:455–460

Colino AF, Arias FJ, Alonso M, Rodriguez-Cabello JC (2014) Self-organized ECM-mimetic model based on an amphiphilic multiblock silk-elastin-like corecombinamer with a concomitant dual physical gelation process. Biomacromolecules 15:3781–3793

Cosgriff-Hernandez E, Hahn MS, Russell B, Wilems T, Munoz-Pinto D, Browning MB, Rivera J, Höök M (2010) Bioactive hydrogels based on designer collagens. Acta Biomater 6:3969–3977

Currie HA, Deschaume O, Naik RR, Perry CC, Kaplan DL (2011) Genetically engineered chimeric silk–silver binding proteins. Adv Funct Mater 21:2889–2895

Dabirian F, Ravandi SAH, Hinestroza JP, Abuzade RA (2012) Conformal coating of yarns and wires with electrospun nanofibers. Polym Eng Sci 52:1724–1732

Davis NR, Anwar RA (1970) Mechanism of formation of desmosine and isodesmosine cross-links of elastin. J Am Chem Soc 92:3778–3782

DiMarco RL, Heilshorn SC (2012) Multifunctional materials through modular protein engineering. Adv Mater 24:3923–3940

Dong H, Zhang D (2014) Current development in genetic engineering strategies of *Bacillus* species. Microb Cell Factories 13:63–74

Dosunmu OO, Chase GG, Kataphinan W, Reneker DH (2006) Electrospinning of polymer nanofibers from multiple jets on a porous tubular surface. Nanotechnology 17:1123–1127

Elliot GF, Huxley AF, Weis-Fogh T (1965) On the structure of resilin. J Mol Biol 13:791–795

Elvin CM, Carr AG, Huson MG, Maxwell JM, Pearson RD, Vuocolo T, Liyou NE, Wong DCC, Merritt DJ, Dixon NE (2005) Synthesis and properties of crosslinked recombinant pro-resilin. Nature 437:999–1002

Fan L, Wang H, Zhang K, He C, Cai Z, Mo X (2012) Regenerated silk fibroin nanofibrous matrices treated with 75 % ethanol vapor for tissue-engineering applications. J Biomater Sci Polym Ed 23:497–508

Ferrer-Miralles N, Villaverde A (2013) Bacterial cell factories for recombinant protein production; expanding the catalogue. Microb Cell Factories 12:113–116

Fleming WW, Sullivan CE, Torchia DA (1980) Characterization of molecular motions in ^{13}C-labeled aortic elastin by ^{13}C-^{1}H magnetic double resonance. Biopolymers 19:597–617

Flory PJ (1953) Principles of polymer chemistry. Cornell University Press, Ithaca/New York

Flower NE, Kenchington W (1967) Studies on insect fibrous proteins: the larval silk of *Apis*, *Bombus* and *Vespa* (Hymenoptera: Aculeata). J Royal Microscop Soc 86:297–310

Foo CWP, Patwardhan SV, Belton DJ, Kitchel B, Anastasiades D, Huang J, Naik RR, Perry CC, Kaplan DL (2006) Novel nanocomposites from spider silk-silica fusion (chimeric) proteins. Proc Natl Acad Sci U S A 103:9428–9433

Fraser RDB, Macrae TP, Millward GR, Parry DAD, Suzuki E, Tulloch PA (1971) The molecular structure of keratins. Appl Polymer Symp 18:65–83

Fu C, Shao Z, Vollrath F (2009) Animal silks: their structures, properties and artificial production. Chem Commun 43:6515–6529

Genç G, Narin G, Bayraktar O (2009) Spray drying as a method of producing silk sericin powders. JAMMR 37:78–86

Gerrard JA, Fayle SE, Wilson AJ, Newberry MP, Ross M, Kavale S (1998) Dough properties and crumb strength of white pan bread as affected by microbial transglutaminase. J Food Sci 63:472–475

Gerrard JA, Brown PK, Fayle SE (2003a) Maillard crosslinking of food proteins ii: the reactions of glutaraldehyde, formaldehyde and glyceraldehyde with wheat proteins *in vitro* and *in situ*. Food Chem 80:35–43

Gerrard JA, Brown PK, Fayle SE (2003b) Maillard crosslinking of food proteins iii: the effects of glutaraldehyde, formaldehyde and glyceraldehyde upon bread and croissants. Food Chem 80:45–50

Gholami A, Tavanai H, Moradi AR (2011) Production of fibroin nanopowder through electrospraying. J Nanopart Res 13:2089–2098

Ghosh S, Parker ST, Wang X, Kaplan DL, Lewis JA (2008) Direct-write assembly of microperiodic silk fibroin scaffolds for tissue engineering applications. Adv Funct Mater 18:1883–1889

Gillespie JM (1990) The Proteins of Hair and Other Hard α-Keratins. In: Goldman R, Steinert P (eds) Cellular and Molecular Biology of Intermediate Filaments. Springer, New York, pp 95–128

Girotti A, Reguera J, Rodriguez-Cabello JC, Arias FJ, Alonso M, Testera AM (2004) Design and bioproduction of a recombinant multi(bio)functional elastin-like protein polymer containing cell adhesion sequences for tissue engineering purposes. J Mater Sci Mater Med 15:479–484

Gnesa E, Hsia Y, Yarger JL, Weber W, Lin-Cereghino J, Lin-Cereghino G, Tang S, Agari K, Vierra C (2012) Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to extensibility in synthetic fibers. Biomacromolecules 13:304–312

Gomes G, Silva-Zacarin EC, Zara FJ, Silva de Moraes RL, Caetano FH (2004) Macromolecular array patterns of silk gland secretion in social Hymenoptera larvae. Genet Mol Res 3:309–322

Gomes SC, Leonor IB, Mano JF, Reis RL, Kaplan DL (2011) Antimicrobial functionalized genetically engineered spider silk. Biomaterials 32:4255–4266

Gorham SD, Light ND, Diamond AM, Willins MJ, Bailey AJ, Wess TJ, Leslie NJ (1992) Effect of chemical modifications on the susceptibility of collagen to proteolysis II Dehydrothermal crosslinking. Int J Biol Macromol 14:129–138

Hacker DL, De Jesus M, Wurm FM (2009) 25 years of recombinant proteins from reactor-grown cells – where do we go from here? Biotechnol Adv 27:1023–1027

Hagn F, Eisoldt L, Hardy JG, Vendrely C, Coles M, Scheibel T, Kessler H (2010) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. Nature 465:239–242

Hardy JG, Scheibel TR (2009) Production and processing of spider silk proteins. J Polym Sci A Polym Chem 47:3957–3963

Hardy JG, Scheibel TR (2010) Composite materials based on silk proteins. Prog Polym Sci 35:1093–1115

Hardy JG, Römer LM, Scheibel TR (2008) Polymeric materials based on silk proteins. Polymer 49:4309–4327

Heikkilä P, Sipilä A, Peltola M, Harlin A (2007) Electrospun PA-66 coating on textile surfaces. Text Res J 77:864–870

Heilshorn SC, Liu JC, Tirrell DA (2005) Cell-binding domain context affects cell behavior on engineered proteins. Biomacromolecules 6:318–323

Hino T, Tanimoto M, Shimabayashi S (2003) Change in secondary structure of silk fibroin during preparation of its microspheres by spray-drying and exposure to humid atmosphere. J Colloid Interface Sci 266:68–73

Hoeve CAJ, Flory PJ (1974) The elastic properties of elastin. Biopolymers 13:677–686

Holland GP, Creager MS, Jenkins JE, Lewis RV, Yarger JL (2008) Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy. J Am Chem Soc 130:9871–9877

Hu X, Kaplan DL, Cebe P (2006) Determining beta-sheet crystallinity in fibrous proteins by thermal analysis and infrared spectroscopy. Macromolecules 39:6161–6170

Hu X, Shmelev K, Sun L, Gil ES, Park SH, Cebe P, Kaplan DL (2011) Regulation of silk material structure by temperature-controlled water vapor annealing. Biomacromolecules 12:1686–1696

Huang J, Foo C, George A, Kaplan DL (2007) The effect of genetically engineered spider silk-dentin matrix protein 1 chimeric protein on hydroxyapatite nucleation. Biomaterials 28:2358–2367

Humenik M, Smith AM, Scheibel T (2011) Recombinant spider silks – biopolymers with potential for future applications. Polymers 3:640–661

Huson MG, Church JS, Poole JM, Weisman S, Sriskantha A, Warden AC, Campbell PM, Ramshaw JAM, Sutherland TD (2012) Controlling the molecular structure and physical properties of artificial honeybee silk by heating or by immersion in solvents. PLoS One 7:e52308

Hwang KY, Yu W-R (2009) Measuring tensile strength of nanofibers using conductive substrates and dynamic mechanical analyser. Fiber Polym 10:703–708

Hwang S, Shao Q, Williams H, Hilty C, Gao YQ (2011) Methanol strengthens hydrogen bonds and weakens hydrophobic interactions in proteins – a combined molecular dynamics and NMR study. J Phys Chem B 115:6653–6660

Jiankang H, Dichen L, Yaxiong L, Bo Y, Bingheng L, Qin L (2007) Fabrication and characterization of chitosan/gelatin porous scaffolds with predefined internal microstructures. Polymer 48:4578–4588

Jin H-J, Kaplan DL (2003) Mechanism of silk processing in insects and spiders. Nature 424:1057–1061

Jones EY, Miller A (1991) Analysis of structural design features in collagen. J Mol Biol 218:209–219

Jose RR, Brown JE, Polido KE, Omenetto FG, Kaplan DL (2015) Polyol-silk bioink formulations as two-part room-temperature curable materials for 3D printing. Biomater Sci Eng 1:780–788

Kameda T (2012) Quantifying the fraction of alanine residues in an α-helical conformation in hornet silk using solid-state NMR. Polymer J 44:876–881

Kameda T, Tamada Y (2009) Variable-temperature ^{13}C solid-state NMR study of the molecular structure of honeybee wax and silk. Int J Biol Macromol 44:64–69

Kameda T, Kojima K, Miyazawa M, Fujiwara S (2005) Film formation and structural characterization of silk of the hornet *Vespa simillima xanthoptera* Cameron. Z Naturforsch 60:906–914

Kameda T, Kojima K, Togawa E, Sezutsu H, Zhang Q, Teramoto H, Tamada Y (2010) Drawing-induced changes in morphology and mechanical properties of hornet silk gel films. Biomacromolecules 11:1009–1018

Kim G, Ahn S, Yoon H, Kim Y, Chun W (2009) A cryogenic direct-plotting system for fabrication of 3D collagen scaffolds for tissue engineering. J Mater Chem 19:8817–8823

Kim MK, Lee JY, Oh H, Song DW, Kwak HW, Yun H, Um IC, Park YH, Lee KH (2015a) Effect of shear viscosity on the preparation of sphere-like silk fibroin microparticles by electrospraying. Int J Biol Macromol 79:988–995

Kim SY, Naskar D, Kundu SC, Bishop DP, Doble PA, Boddy AV, Chan H-K, Wall IB, Chrzanowski W (2015b) Formulation of biologically-inspired silk-based drug carriers for pulmonary delivery targeted for lung cancer. Sci Rep 5:11878

Koria P, Yagi H, Kitagawa Y, Megeed Z, Nahmias Y, Sheridan R, Yarmush ML (2011) Self-assembling elastin-like peptides growth factor chimeric nanoparticles for the treatment of chronic wounds. Proc Natl Acad Sci U S A 108:1034–1039

Krishnaji ST, Kaplan DL (2013) Bioengineered chimerics silk-uranium binding proteins. Macromol Biosci 13:256–264

Lai T, Yang Y, Ng SK (2013) Advances in mammalian cell line development technologies for recombinant protein production. Pharmaceuticals 6:579–603

Lammel A, Schwab M, Slotta U, Winter G, Scheibel T (2008) Processing conditions for the formation of spider silk microspheres. Chem Sus Chem 1:413–416

Lawrence BD, Wharram S, Kluge JA, Leisk GG, Omenetto FG, Rosenblatt MI, Kaplan DL (2002) Effect of hydration on silk film material properties. Macromol Biosci 10:393–403

Lawrence BD, Omenetto F, Chui K, Kaplan DL (2008) Processing methods to control silk fibroin film biomaterial features. J Mater Sci 43:6967–6985

Lefèvre T, Rousseau M-E, Pézolet M (2007) Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. Biophys J 92:2885–2895

Li B, Daggett V (2002) The molecular basis of the temperature- and pH-induced conformational transitions in elastin-based peptides. Biopolymers 68:121–129

Li Z, Wang C (2013) Effects of working parameters on electrospinning. In: One dimensional nanostructures. Electrospinning technique and unique nanofibers. Springer, Berlin/Heidelberg, pp 15–28

Lin S, Ryu S, Tokareva O, Gronau G, Jacobsen MM, Huang W, Rizzo DJ, Li D, Staii C, Pugno NM, Wong JY, Kaplan DL, Buehler MJ (2015) Predictive modelling-based design and experiments for synthesis and spinning of bioinspired silk fibres. Nat Commun 6:6892–7904

Liu JC, Tirrell DA (2008) Cell response to RGD density in cross-linked artificial extracellular matrix protein films. Biomacromolecules 9:2984–2988

Liu JC, Heilshorn SC, Tirrell DA (2004) Comparative cell response to artificial extracellular matrix proteins containing the RGD and CS5 cell-binding domains. Biomacromolecules 5:497–504

Lu Q, Hu X, Wang X, Kluge JA, Lu S, Cebe P, Kaplan DL (2010) Water-insoluble silk films with silk I structure. Acta Biomater 6:1380–1387

Lupas AN, Gruber M (2005) The structure of α-helical coiled coils. Adv Protein Chem 70:37–78

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

Lyerla JR Jr, Torchia DA (1975) Molecular mobility and structure of elastin deduced from the solvent and temperature dependence of 13C magnetic resonance relaxation data. Biochemistry 14:5175–5183

Magalhäes PO, Lopes AM, Mazzola PG, Rangel-Yagui C, Penna TCV, Pessoa A Jr (2007) Methods of endotoxin removal from biological preparations: a review. J Pharm Pharmaceut Sci 10:388–404

Malencik DA, Anderson SR (1996) Dityrosine formation in calmodulin: cross-linking and polymerization catalyzed by *Arthromyces* peroxidase. Biochemistry 35:4375–4386

Mamat U, Wilke K, Bramhill D, Schromm AB, Lindner B, Kohl TA, Corchero JL, Villaverde A, Schaffer L, Head SR, Souvignier C, Meredith TC, Woodard RW (2015) Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins. Microb Cell Factories 14:57–71

Marsano E, Corsini P, Arosio C, Boschi A, Mormino M, Freddi G (2005) Wet spinning of *Bombyx mori* silk fibroin dissolved in *N*-methyl morpholine *N*-oxide and properties of regenerated fibres. Int J Biol Macromol 37:179–188

Marshall RC, Orwin DFG, Gillespie JM (1991) Structure and biochemistry of mammalian hard keratin. Electron Microsc Rev:4:47–4:83

Martínez JL, Liu L, Petranovic D, Nielsen J (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. Curr Opin Biotechnol 23:965–971

Massodi I, Bidwell GL III, Raucher D (2005) Evaluation of cell penetrating peptides fused to elastin-like polypeptide for drug delivery. J Control Release 108:396–408

Massodi I, Moktan S, Rawat A, Bidwell GL III, Raucher D (2010) Inhibition of ovarian cancer cell proliferation by a cell cycle inhibitory peptide fused to a thermally responsive polypeptide carrier. Int J Cancer 126:533–544

Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M et al (2012) Recombinant protein production in yeasts. In: Lorence A (ed) Recombinant gene expression. Humana Press, Totowa, pp 329–358

Mieszawska AJ, Nadkarni LD, Perry CC, Kaplan DL (2010) Nanoscale control of silica particle formation via silk−silica fusion proteins for bone regeneration. Chem Mater 22:5780–5785

Mithieux SM, Wise SG, Weiss AS (2013) Tropoelastin – a multifaceted naturally smart material. Adv Drug Deliv Rev 65:421–428

Mo C, Wu P, Chen X, Shao Z (2006) Near-infrared characterization on the secondary structure of regenerated *Bombyx mori* silk fibroin. Appl Spectrosc 60:1438–1441

Mohs A, Silva T, Yoshida T, Amin R, Lukomski S, Inouye M, Brodsky B (2007) Mechanism of stabilization of bacterial collagen triple helix in the absence of hydroxyproline. J Biol Chem 282:29757–29765

Mouw JK, Ou G, Weaver VM (2014) Extracellular matrix assembly: a multiscale deconstruction. Nat Rev Mol Cell Biol 15:771–785

Muiznieks LD, Weiss AS, Keeley FW (2010) Structural disorder and dynamics of elastin. Biochem Cell Biol 88:239–250

Nakamura M, Mie M, Mihara H, Nakamura M, Kobatake E (2009) Construction of a multifunctional extracellular matrix protein that increases number of N1E-115 neuroblast cells having neurites. J Biomed Mater Res B Appl Biomater 91:425–432

Nazarov R, Jin H-J, Kaplan DL (2004) Porous 3-D scaffolds from regenerated silk fibroin. Biomacromolecules 5:718–726

Park SH, Gil ES, Kim HJ, Lee K, Kaplan DL (2010) Relationships between degradability of silk scaffolds and osteogenesis. Biomaterials 31:6162–6172

Partridge SM, Elsden DF, Thomas J, Dorfman A, Telser A, Ho P-L (1966) Incorporation of labelled lysine into the desmosine cross-bridges in elastin. Nature 209:399–400

Peng YY, Howell L, Stoichevska V, Werkmeister JA, Dumsday GJ, Ramshaw JAM (2012) Towards scalable production of a collagen-like protein from *Streptococcus pyogenes* for biomedical applications. Microb Cell Factories 11:146–153

Peng YY, Stoichevska V, Schacht K, Werkmeister JA, Ramshaw JA (2014a) Engineering multiple biological functional motifs into a blank collagen-like protein template from *Streptococcus pyogenes*. J Biomed Mater Res A 102:2189–2196

Peng YY, Stoichevska V, Madsen S, Howell L, Dumsday GJ, Werkmeister JA, Ramshaw JAM (2014b) A simple cost-effective methodology for large-scale purification of recombinant non-animal collagens. Appl Microbiol Biotechnol 98:1807–1815

Pérez-Rigueiro JP, Biancotto L, Corsini P, Marsano E, Elices M, Plaza GR, Guinea GV (2009) Supramolecular organization of regenerated silkworm silk fibers. Int J Biol Macromol 44:195–202

Persikov AV, Ramshaw JAM, Brodsky B (2005) Prediction of collagen stability from amino acid sequence. J Biol Chem 280:19343–19349

Phelps JP, Dang N, Rasochova L (2007) Inactivation and purification of cowpea mosaic virus-like particles displaying peptide antigens from *Bacillus anthracis*. J Virol Methods 141:146–153

Pometun MS, Chekmenev EY, Wittebort RJ (2004) Quantitative observation of backbone disorder in native elastin. J Biol Chem 279:7982–7987

Poole J, Church JS, Woodhead AL, Huson MG, Sriskantha A, Kyratzis IL, Sutherland TD (2013) Continuous production of flexible fibers from transgenically produced honeybee silk proteins. Macromol Biosci 13:1321–1326

Qin G, Hu X, Peggy Cebe P, Kaplan DL (2012) Mechanism of resilin elasticity. Nat Commun 3:1003

Que RA, Chan SWP, Jabaiah AM, Lathrop RH, Da Silva NA, Wang SW (2015) Tuning cellular response by modular design of bioactive domains in collagen. Biomaterials 53:309–317

Rajkhowa R, Levin B, Redmond SL, Li LH, Wang L, Kanwar JR, Atlas MD, Wang X (2011) Structure and properties of biomedical films prepared from aqueous and acidic silk fibroin solutions. J Biomed Mater Res A 97:37–45

Ramachandran G, Kartha G (1954) Structure of collagen. Nature 174:269–270

Rapson TD, Church JS, Trueman HE, Dacres H, Sutherland TD, Trowell SC (2014) Micromolar biosensing of nitric oxide using myoglobin immobilised in a synthetic silk film. Biosens Bioelectron 62:214–220

Rapson TD, Sutherland TD, Church JS, Trueman HE, Dacres H, Trowell SC (2015) *de novo* engineering of solid-state metalloproteins using recombinant coiled-coil silk. Biomater Sci Eng 1:1114–1120

Rasmussen M, Jacobsson M, Björck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. J Biol Chem 278:32313–32316

Reiser K, McCormick RJ, Rucker RB (1992) Enzymatic and nonenzymatic cross-linking of collagen and elastin. FASEB J 6:2439–2449

Rockwood DN, Preda RC, Yücel T, Wang X, Lovett ML, Kaplan DL (2011) Materials fabrication from *Bombyx mori* silk fibroin. Nat Protoc 6:1612–1631

Rodriguez-Cabello JC, Reguera J, Girotti A, Arias J, Alonso M (2006) Genetic engineering of protein-based polymers: the example of elastinlike polymers. In Vancso GJ (ed) Ordered Polymeric Nanostructures at Surfaces Adv Polym Sci 200, Springer, Berlin, p 119–167

Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol 5:1–17

Rudall KM (1962) Silk and other cocoon proteins. In: Florkin M, Mason HS (eds) Comparative Biochemistry 4: a comprehensive treatise. Academic Press, New York, pp 397–433

Schallmey M, Singh A, Ward OP (2004) Developments in the use of *Bacillus* species for industrial production. Can J Microbiol 50:1–17

Schumann W (2007) Production of recombinant proteins in *Bacillus subtilis*. Adv Appl Microbiol 62:137–189

Seidel A, Liivak O, Jelinski LW (1998) Artificial spinning of spider silk. Macromolecules 31:6733–6736

Seidel A, Liivak O, Calve S, Adaska J, Ji GD, Yang ZT, Grubb D, Zax DB, Jelinski LW (2000) Regenerated spider silk: processing, properties and structure. Macromolecules 33:775–780

Seo N, Russell BH, Rivera JJ, Liang X, Xu X, Afshar-Kharghan V, Höök M (2010) An engineered α1 integrin-binding collagenous sequence. J Biol Chem 285:31046–31054

Sezutsu H, Kajiwara H, Kojima K, Mita K, Tamura T, Tamada Y, Kameda T (2007) Identification of four major hornet silk genes with a complex of alanine-rich and serine-rich sequences in *Vespa simillima xanthoptera* Cameron. Biosci Biotechnol Biochem 71:2725–2734

Shao ZZ, Vollrath F, Yang Y, Thøgersen HC (2003) Structure and behavior of regenerated spider silk. Macromolecules 36:1157–1161

Shi J, Lua S, Du N, Liu X, Song J (2008) Identification, recombinant production and structural characterization of four silk proteins from the Asiatic honeybee *Apis cerana*. Biomaterials 29:2820–2828

Siegel RC, Pinnell SR, Martin GR (1970) Cross-linking of collagen and elastin. Properties of lysyl oxidase. Biochemist 9:4486–4492

Silva-Zacarin ECM, Silva de Moraes RLM, Taboga SR (2003) Silk formation mechanisms in the larval salivary glands of *Apis mellifera* (Hymenoptera:Apidae). J Biosci 28:753–764

Slotta UK, Rammensee S, Gorb S, Scheibel T (2008) An engineered spider silk protein forms microspheres. Angew Chem 47:4592–4594

Straley KS, Heilshorn SC (2009) Independent tuning of multiple biomaterial properties using protein engineering. Soft Matter 5:114–124

Sutherland TD, Weisman S, Trueman TE, Sriskantha A, Trueman JWH, Haritos VS (2007) Conservation of essential design features in coiled coil silks. Mol Biol Evol 24:2424–2432

Sutherland TD, Young J, Weisman S, Hayashi CY, Merrit D (2010) Insect silk: one name, many materials. Annu Rev Entomol 55:171–188

Sutherland TD, Church JS, Hu X, Huson MG, Kaplan DL, Weisman S (2011) Single honeybee silk protein mimics properties of multi-protein silk. PLoS One 6:e16489

Sutherland TD, Peng YY, Trueman HE, Weisman S, Okada S, Walker AA, Sriskantha A, White JF, Huson MG, Werkmeister JA, Glattauer V, Stoichevska V, Mudie ST, Haritos VS, Ramshaw JAM (2013) A new class of animal collagen masquerading as an insect silk. Sci Rep 3:2864

Sutherland TD, Trueman HE, Walker AA, Weisman S, Campbell PM, Dong Z, Huson MG, Woodhead AL, Church JS (2014) Convergently-evolved structural anomalies in the coiled coil domains of insect silk proteins. J Struct Biol 186:402–411

Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, Forlino A, Cabral WA, Barnes AM, Marini JC, San Antonio JD (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J Biol Chem 283:21187–21197

Takagi H, Kadowaki K, Udaka S (1989) Screening and characterization of protein-hyperproducing bacteria without detectable exoprotease activity. Agric Biol Chem 53:691–699

Teulé F, Miao Y-G, Sohn B-H, Kim Y-S, Hull JJ, Fraser MJ Jr, Lewis RV, Jarvis DL (2012) Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties. Proc Natl Acad Sci U S A 109:923–928

Tjin MS, Low P, Fong E (2014) Recombinant elastomeric protein biopolymers: progress and prospects. Polymer J 46:444–451

Tokareva O, Michalczechen-Lacerda VA, Rech EL, Kaplan DL (2013) Recombinant DNA production of spider silk proteins. Microb Biotechnol 6:651–663

Tokareva O, Jacobsen M, Buehler M, Wong J, Kaplan DL (2014) Structure-function-property-design interplay in biopolymers: spider silk. Acta Biomater 10:1612–1626

Torchia DA, Piez KA (1973) Mobility of elastin chains as determined by ^{13}C nuclear magnetic resonance. J Mol Biol 76:419–424

Tsukada M, Gotoh Y, Nagura M, Minoura N, Kasai N, Freddi G (1994) Structural changes of silk fibroin membranes induced by immersion in methanol aqueous solutions. J Polym Sci Part B 32:961–968

Urry DW (1988) Entropic elastic processes in protein mechanisms I Elastic structure due to an inverse temperature transition and elasticity due to internal chain dynamics. J Protein Chem 7:1–34

Urry D, Luan C-H, Peng SQ (1995) Molecular biophysics of elastin structure, function and pathology. Ciba Foundation Symp 192:4–30

van Beek JD, Beaulieu L, Schäfer H, Demura M, Asakura T, Meier BH (2000) Solid-state NMR determination of the secondary structure of *Samia cynthia ricini* silk. Nature 405:1077–1079

van Eldijk MB, McGann CL, Kiick KL, van Hest JCM (2012) Elastomeric polypeptides. Top Curr Chem 310:71–116

van Vlierberghe S, Dubruel P, Schacht E (2011) Biopolymer-based hydrogels as scaffolds for tissue engineering applications: a review. Biomacromolecules 12:1387–1408

Varabhas JS, Chase GG, Reneker DH (2008) Electrospun nanofibers from a porous hollow tube. Polymer 49:4226–4229

Vepari C, Kaplan DL (2007) Silk as a biomaterial. Prog Polym Sci 32:991–1007

Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. Nature 410:541–548

Vollrath F, Porter D (2006) Spider silk as archetypal protein elastomer. Soft Matter 2:377–385

Walker AA, Weisman S, Church JS, Merritt DJ, Mudie ST, Sutherland TD (2012) Silk from crickets: a new twist on spinning. PLoS One 7:e30408

Walker AA, Church JS, Woodhead AL, Sutherland TD (2013a) Silverfish silk is formed by entanglement of randomly coiled protein chains. Insect Biochem Mol Biol 43:572–579

Walker AA, Warden AC, Trueman HE, Weisman S, Sutherland TD (2013b) Micellar refolding of coiled-coil honeybee silk proteins. J Mater Chem B 1:3644–3651

Walker AA, Holland C, Sutherland TD (2015) More than one way to spin a crystallite: multiple trajectories through liquid crystallinity to solid silk. Proc R Soc B 282:20150259

Walsh G (2010) Post-translational modifications of protein biopharmaceuticals. Drug Discov Today 15:773–780

Wang C, Patwardhan SV, Belton DJ, Kitchel B, Anastasiades D, Huang J et al (2006) Novel nanocomposites from spider silk-silica fusion (chimeric) proteins. Proc Natl Acad Sci USA 103:9428–9433

Wang X, Wenk E, Matsumoto A, Meinel L, Li C, Kaplan DL (2007) Silk microspheres for encapsulation and controlled release. J Control Release 117:360–370

Wang X, Kluge JA, Leisk GG, Kaplan DL (2008) Sonication-induced gelation of silk fibroin for cell encapsulation. Biomaterials 29:1054–1064

Wang X, Yucel T, Lu Q, Hu X, Kapla DL (2010) Silk nanospheres and microspheres from silk/pva blend films for drug delivery. Biomaterials 31:1025–1035

Wang X, Niu H, Wang X, Lin T (2012) Needleless electrospinning of uniform nanofibers using spiral coil spinnerets. J Nanomat 2012:785920

Weis-Fogh T (1961a) Thermodynamic properties of resilin, a rubber-like protein. J Mol Biol 3:520–531

Weis-Fogh T (1961b) Molecular interpretation of the elasticity of resilin, a rubber-like protein. J Mol Biol 3:648–667

Weisman S, Haritos VS, Church JS, Huson MG, Mudie ST, Rodgers AJW, Dumsday GJ, Sutherland TD (2010) Honeybee silk: recombinant protein production, assembly and fiber spinning. Biomaterials 31:2695–2700

Werkmeister JA, Ramshaw JAM (2012) Recombinant protein scaffolds for tissue engineering. Biomed Mater 7:012002

Widhe M, Johansson U, Hillerdahl C-O, Hedhammar M (2013) Recombinant spider silk with cell binding motifs for specific adherence of cells. Biomaterials 34:8223–8234

Wohlrab S, Müller S, Schmidt A, Neubauer S, Kessler H, Leal-Egaña A, Scheibel T (2012) Cell adhesion and proliferation on RGD-modified recombinant spider silk proteins. Biomaterials 33:6650–6659

Xia XX, Qian Z-G, Ki CS, Park YH, Kaplan DL, Lee SY (2010) Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. Proc Natl Acad Sci U S A 107:14059–14063

Xia XX, Xu Q, Hu X, Qin G, Kaplan DL (2011) Tunable self-assembly of genetically engineered silk-elastin-like protein polymers. Biomacromolecules 12:3844–3850

Xie F, Zhang H, Shao H, Hu X (2006) Effect of shearing on formation of silk fibers from regenerated *Bombyx mori* silk fibroin aqueous solution. Int J Biol Macromol 38:284–288

Yan J, Zhou G, Knight DP, Shao Z, Chen X (2010) Wet-spinning of regenerated silk fiber from aqueous silk fibroin solution: discussion of spinning parameters. Biomacromolecules 11:1–5

Yannas IV, Tobolsky AV (1967) Cross-linking of gelatin by dehydration. Nature 215:509–510

Yener F, Jirsak O (2012) Comparison between the needle and roller electrospinning of polyvinyl-butyral. J Nanomater 2012:839317

Yu Z, An B, Ramshaw JAM, Brodsky B (2014) Bacterial collagen-like proteins that form triple helical structures. J Struct Biol 186:451–461

Zara JF, Caetano FH (2002) Ultrastructure of the salivary glands of *Pachycondyla (= Neoponera) villosa* (Fabricius) (Formicidae: Ponerinae): functional changes during the last larval instar. Cytologia 67:267–280

Zhang Y, Malamakal RM, Chenoweth DM (2015) Aza-glycine induces collagen hyperstability. J Am Chem Soc 137:12422–12425

Zhao X, Fox JM, Olson NH, Baker TS, Young MJ (1995) *In vitro* assembly of cowpea chlorotic mottle virus from coat protein expressed in *Esherichia coli* and *in vitro*-transcribed viral cDNA. Virology 207:486–494

Zhou G, Shao Z, Knight DP, Yan J, Chen X (2009) Silk fibers extruded artificially from aqueous solutions of regenerated *Bombyx mori* silk fibroin are tougher than their natural counterparts. Adv Mater 21:366–370

Zhou H, Shao Z, Chen X (2014) Wet-spinning of regenerated silk fiber from aqueous silk fibroin solutions: influence of calcium ion addition in spinning dope on the performance of regenerated silk fiber. Chin J Polym Sci 32:29–34

Ziegler K (1977) Crosslinking and self-crosslinking in keratin fibres. In: Asquith RS (ed) Chemistry of natural protein fibers. Springer, New York, pp 267–300

Chapter 16
Properties of Engineered and Fabricated Silks

Gregor Lang, Heike Herold, and Thomas Scheibel

Contents

Abstract Silk is a protein-based material which is predominantly produced by insects and spiders. Hundreds of millions of years of evolution have enabled these animals to utilize different, highly adapted silk types in a broad variety of applications. Silk occurs in several morphologies, such as sticky glue or in the shape of fibers and can, depending on the application by the respective animal, dissipate a high mechanical energy, resist heat and radiation, maintain functionality when submerged in water and withstand microbial settling. Hence, it's unsurprising that silk

G. Lang
Research Group Biopolymer Processing, University of Bayreuth,
Universitätsstr. 30, 95440 Bayreuth, Germany

H. Herold • T. Scheibel (✉)
Department of Biomaterials, University of Bayreuth,
Universitätsstr. 30, 95440 Bayreuth, Germany
e-mail: thomas.scheibel@bm.uni-bayreuth.de

piqued human interest a long time ago, which catalyzed the domestication of silkworms for the production of silk to be used in textiles. Recently, scientific progress has enabled the development of analytic tools to gain profound insights into the characteristics of silk proteins. Based on these investigations, the biotechnological production of artificial and engineered silk has been accomplished, which allows the production of a sufficient amount of silk materials for several industrial applications. This chapter provides a review on the biotechnological production of various silk proteins from different species, as well as on the processing techniques to fabricate application-oriented material morphologies.

Keywords Silk fibers • Artificial silk • Recombinant silk • Engineered silk • Silk processing

16.1 Introduction

Silk is a highly fascinating natural material. It has inspired humans throughout history and has been exploited for many purposes. Precious cloths of silky shine have been produced from the silk of the silkworm *Bombyx mori* for millennia. Domestication of the silk moth over 5000 years ago together with improvements in processing of raw silk, e.g. removal of the glue proteins coating the natural silk fiber, gave rise to a huge and economically important textile industry (Sun et al. 2012). Silks have further attracted much attention, as some silk fibers exhibit extraordinary properties, such as great strength and extensibility, resulting in a great toughness in combination with good biocompatibility, making silks promising candidates for many applications in the biomedical field, but also a favorable material for technical applications (Vollrath 2000; Vollrath and Knight 2001).

Ancient Romans and Greeks already made use of the low immunogenicity of silk and used spider silk webs to stop bleedings and to cover wounds (Scheibel 2004). As a result of biocompatibility and strength, silk has also been used as suture material. The extraordinary mechanical properties of silk have further been exploited to string branches with spider silk in order to create a tear resistant dip net for fishing or for the construction of thin and therefore precise as well as durable crosshairs for the military (Heim et al. 2009). Even violin strings have been produced from spider silk (Osaki 2012). Due to their enormous diversity and due to the fact that all silks have been the object of an evolutionary optimization process, they might provide a basis for the development of a plethora of future materials. Great efforts have been made to investigate silks, to produce the underlying proteins recombinantly and to process the proteins into different morphologies to exploit the beneficial properties of silks in various applications (Fig. 16.1).

Silks have evolved several times among arthropods and are as such a generally diverse material. They function as protein-based extracorporeal structural materials which are produced by a wide range of insects and spiders for numerous purposes, e.g. to capture prey, for adhesion, to build cocoons for pupation or to serve as egg stalks to just name a few examples (Craig 1997; Sutherland et al. 2010). Generally,
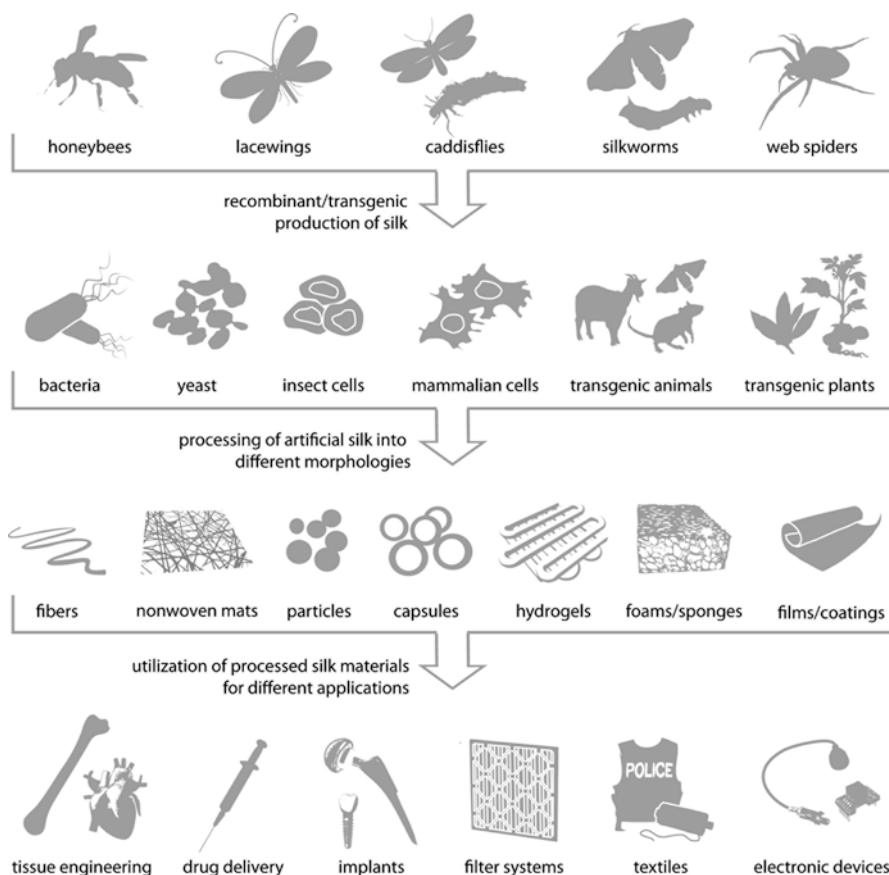
Fig. 16.1 Overview of natural silk producers, host organisms for the production of recombinant silk, range of feasible morphologies and potential applications

silk proteins are secreted into a silk gland and stored there as a highly concentrated spinning dope, which is further treated in a fiber spinning process to create insoluble fibers from the liquid dope (Neuenfeldt and Scheibel 2014). The best studied silks are cocoon silk produced by caterpillars of silk moths and webs of orb-weaving spiders (Zhou et al. 2001; Vepari and Kaplan 2007; Heim et al. 2010). Despite their evolutionary distance, the silk proteins of insects and spiders share several properties: they are often large proteins with molecular weights of over 150 kDa, they possess a repetitive core domain, where long hydrophobic blocks are alternated with short hydrophilic regions, and the core domain is flanked by hydrophilic, non-repetitive termini (Bini et al. 2004). The amino acid composition of silk proteins reveals a high level of alanine, serine and glycine residues (Garb et al. 2010; Lintz and Scheibel 2013) and accounts for both the stability of the spinning dope as well as for the defined secondary structure upon transition of the solution into an insoluble thread, which in large parts determines the mechanical properties of the fiber.

Table 16.1 Comparison of mechanical properties of silk fibers and synthetic fibers

Material	Strength [GPa]	Extensibility [%]	Toughness [MJ/m³]
honey bee silk (*A. mellifera*)	0.132	204	n.d.
lacewing egg stalk silk *(C. carnea)* (70 % rel. Humidity)	0.155	210	87
caddisfly silk	0.033	126	17
silkworm silk *(B. mori)*	0.6	18	70
spider silk – dragline *(A. diadematus)*	1.1	27	160
nylon fiber	0.95	18	80
kevlar 49 fiber	3.6	2.7	50
carbon fiber	4	1.3	25
high-tensile steel	1.5	0.8	6

Data taken from Gosline et al. (1999), except for data on honey bee silk Hepburn et al. (1979), lacewing silk Bauer and Scheibel (2012) and caddisfly silk Lane et al. (2015)
n.d.: not determined

But, since the diversity is high, other silk concepts have also evolved. The larvae of honeybees produce a silk that consists of much smaller (~30 kDa) and less repetitive silk proteins, which are assembled into α-helical coiled coil structures, in the rather hydrophobic waxen environments of the beehive (Sutherland et al. 2006). Lacewing egg stalks are another example admired for their bending stiffness, which is caused by a cross-β structure (Weisman et al. 2009). In contrast to terrestrial silks, the larvae of caddisflies, as a representative of aquatic silk producers, have developed an underwater-adhesive silk to adhere sand grains and branchlets to their body or their retreat tube (Ashton et al. 2013). The enormous diversity in silks leads to the different mechanical properties of the different silk fibers (Table 16.1).

Recombinant protein production is one key technology towards the employment of silk materials (other than silkworm silk) on an industrial scale. The natural availability of silk proteins – with the exception of *Bombyx mori* silk – is fairly low, and, if harvested from the respective organisms, proteins vary in molecular weight, amount and quality. For instance, attempts to farm spiders failed due to their territorial and cannibalistic behavior (Vendrely and Scheibel 2007). Instead, recombinant production of silk proteins allows silk protein production at large scale with constant quality and furthermore facilitates genetic engineering of the coding gene sequence, allowing the production of tailored proteins for respective applications (Vendrely and Scheibel 2007). Due to the great availability of different engineered strains, as well as the huge number of suitable expression plasmids, *E. coli* is by far the most used host organism for heterologous protein production (Rosano and Ceccarelli 2014). Nevertheless, during the production of (large) silk proteins in bacterial hosts, problems such as low yield, truncated proteins, and aggregation of proteins in inclusion bodies have been encountered (Vendrely and Scheibel 2007; Scheibel 2004; Heidebrecht and Scheibel 2013). To circumvent these obstacles, the nucleotide sequence encoding silk proteins can be engineered with regard to the host's codon usage or limitations in the size of expressible genes. For the expression

of original (partial) silk gene sequences, expression hosts other than *E. coli*, e.g. yeast, insect cells, mammalian cell lines, even transgenic animals and plants have been tested (Vendrely and Scheibel 2007).

As well as silk protein production and protein purification from the expression host, another key step is the processing of the obtained recombinant silk proteins. Besides fibers, which are typically found in nature, many other morphologies, e.g. nonwoven mats, particles, capsules, hydrogels, foams or coatings, are feasible and, therefore, allow diverse applications of silk materials. Fibers spun into nonwoven mats can improve filter systems by an excellent filter efficiency and much better air permeability than comparable materials (Lang et al. 2013). The excellent mechanical properties of spider silk fibers could improve textiles such as bullet-proof vests or other textiles where extensibility combined with tensile strength and light weight are in demand. Even electronic devices could benefit from elaborated silk protein processing. Silks cannot only serve as a passive substrate in bio-integrated electronics, but can also be used as active elements, e.g. as transistors or memristive devices (Zhu et al. 2015). Due to their good biocompatibility, silks have many applications in biomedicine, e.g. in tissue engineering. Cell-seeded three-dimensional constructs have good prospects for *in vitro*-cultured replacements of tissues, in which the biodegradable silk scaffold degrades at the rate the body accepts and integrates the material. Biodegradable silk particles or capsules can be used as mobile drug carriers, whereas other morphologies mentioned above can be employed as stationary drug delivery platforms, e.g. to achieve constant drug levels over time through sustained drug release (Blüm and Scheibel 2012). Silk-coatings of implants can improve the biocompatibility of the material and thereby reduce unwanted reactions like fibrosis (Zeplin et al. 2014b; Borkner et al. 2014).

16.2 Biotechnological Production of Silk

For the recombinant production of proteins, usually the desired genetic information encoding the protein is cloned into an appropriate vector from which the gene can be expressed in a suitable expression host. Nevertheless, for many silk proteins, the gene encoding the silk protein is unknown or not fully known and, therefore, has to be identified prior to recombinant production. As an example, the identification of four honeybee silk genes is depicted here. The silk gene expression is highly regulated in honey bee. Since only final instar larvae produce silk in their modified salivary glands (Sutherland et al. 2006), samples were taken from these glands and analyzed with regard to gene expression, i.e. mRNAs were analyzed by creation of a cDNA library (Fig. 16.2). As RNA is prone to degradation, reverse transcriptase is used to transcribe mRNA into cDNA. As most eukaryotic RNA is poly-adenylated, the poly(A) tails of the 3′ ends allow easy transcription by use of oligo(dT) primers. The RNA moiety can then be degraded by RNase H, and after a PCR-amplification cycle double-stranded cDNA emerges. Terminal transferase subsequently adds
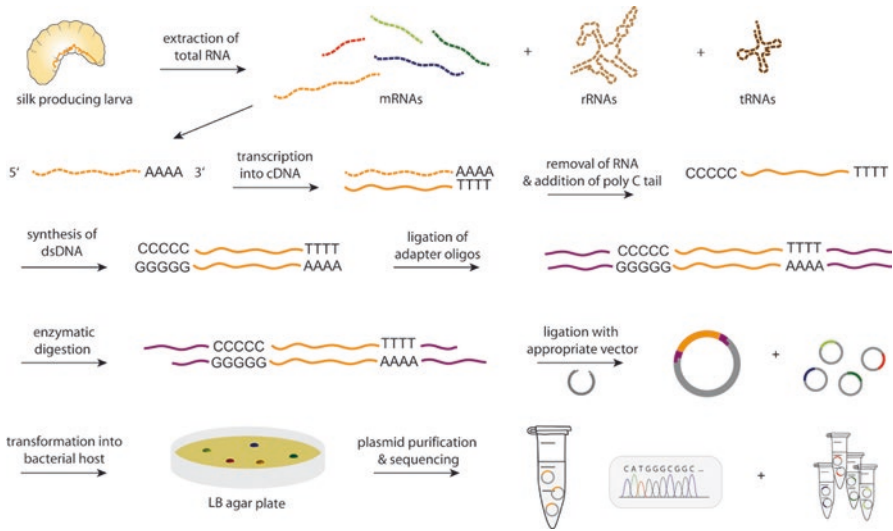
Fig. 16.2 Creation of a cDNA library. Extracted mRNAs are transcribed into cDNA and finally cloned into a vector to allow sequencing of the plasmids upon amplification in bacteria, e.g. *E. coli*

several Cs to the terminal parts of the cDNA, which allow poly(G) primers to be used for the amplification of the constructs. Cohesive ends can then be added to the DNA strands, enabling ligation into a plasmid with complementary ends. This enables efficient sequencing of the unknown nucleotide sequences as the sequence of the vector is well-known and, therefore, allows precise design of sequencing primers. The ligated vectors are transformed into suitable expression hosts, e.g. *E. coli*, supporting plasmid propagation and facilitating plasmid preparation at reasonable quantities. As this process is carried out with all mRNAs present in the sample, a so-called cDNA library emerges. Thus, the cDNA library represents the population of mRNAs present in the cell at a certain time point and, remarkably, only contains exons of a gene.

In the case of the honey bee silk proteins, the proteins present in the gland were purified and analyzed by mass spectrometry after tryptic digestion. These fragments were matched with the predicted protein sizes (after tryptic digestion) of the transcriptome obtained from the generated cDNA library of the analyzed silk gland tissue and further matched with predictions obtained from genomic information. Taken together, six silk-encoding genes could be identified. Further analysis of the encoded proteins revealed that four genes encoded fibrous silk proteins of 30–34 kDa (termed *Apis mellifera* fibroin 1–4), whereas the remaining two proteins were classified as silk-associated proteins, with one small gene encoding a possible glue protein and the second one a high molecular weight protein with a yet unknown function (Sutherland et al. 2006).
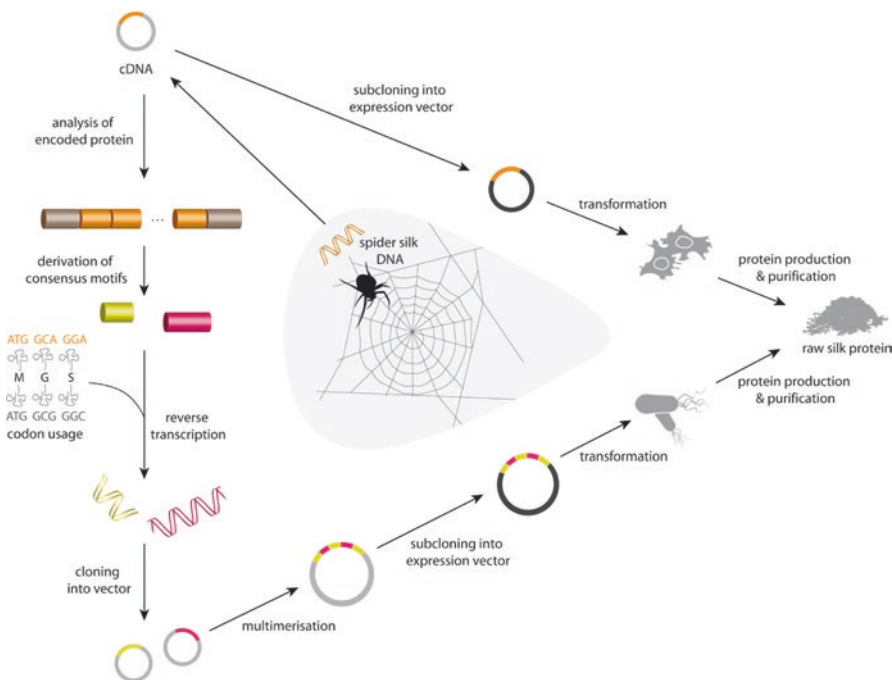
Fig. 16.3 Cloning strategies for recombinant silk production. For the production of original (partial) silk fragments, appropriate hosts have to be chosen, while for the production of large silk proteins in bacterial hosts such as *E. coli* synthetic silk genes have to be designed to facilitate reasonable protein production rates

Once a silk gene is identified it is cloned into an expression vector and transformed into the respective expression host to produce the recombinant protein. While this works perfectly well for smaller proteins with less repetitive sequences, such as honey bee silk proteins, the production of larger proteins with highly repetitive sequences, such as silkworm silk or spider silk proteins, often suffers from low yields of recombinant (full-length) protein. The underlying silk genes can have a size of up to 15 Mbp which conflicts with the limited size of well-expressible genes in e.g. the preferred bacterial expression host *E. coli*. Premature termination of the transcriptional and/or translational process together with depletion of specific tRNAs caused by differences in codon usage of the expression host often lead to fragmented silk proteins. Furthermore, the repetitive sequences can undergo homologous recombination, thereby shortening the genetic information (Vendrely and Scheibel 2007; Arcidiacono et al. 1998). Therefore, either a suitable expression host has to be chosen for the expression of original silk genes or synthetic silk-like genes with adopted codon usage and gene length have to be created (Fig. 16.3).

16.2.1 Examples of Biotechnologically Produced Silk Proteins: Honey Bee Silk

16.2.1.1 Natural Honey Bee Silk

Honey bees produce silk in their larval stage (Sutherland et al. 2006). Like many other holometabolous insects, the final instar larvae of honey bees of the species *Apis mellifera* produce a silken cocoon to undergo metamorphosis inside their sealed brood comb cells (Sutherland et al. 2006). The deposition of silk has beneficial effects on the hive in general as it increases its thermal stability and its mechanical strength, probably by creating a structure similar to a modern fiber-reinforced composite (Zhang et al. 2010; Weisman et al. 2010). The larvae attach a silken layer to the waxen walls of the brood comb cells prior to pupation, and from one generation to the next the proportion of silk accumulates and can increase up to 40 % of the total nest mass (Hepburn and Kurstjens 1988). Mechanical tests revealed that old combs are much stronger and stiffer, and especially at temperatures above 40 °C the composite material of silk and wax is up to one order of magnitude stronger than wax alone (Hepburn and Kurstjens 1988; Zhang et al. 2010). Little is known about the mechanical properties of the plain larval silks of the Vespoidea and Apoidea, the latter including honey bee silk, due to the difficulties in obtaining appropriate material from living organisms (Sutherland et al. 2012). Hand-drawn fibers of honey bee larvae revealed a breaking stress of 132 MPa and a breaking strain of 204 % at a relative humidity of 65 % (Hepburn et al. 1979). They have a greater extensibility and toughness compared to silkworm silk but lower tensile strength (Hepburn et al. 1979). Honeybee silk is a very interesting material, as it is very extensible and also tough and keeps its properties when wet (Hepburn et al. 1979).

Honey bee larvae produce silk from modified salivary glands (labial glands), a feature they share with silkworm silk (Sutherland et al. 2006). The silk forming mechanism of *Apis mellifera* in the labial glands was reported to be similar to that of Lepidoptera. Within the gland, the silk proteins assemble into cigar-shaped bodies (Weisman et al. 2010). Subsequent to polymerization of the proteins in the lumen of the gland into so-called tactoids, a gradual dehydration during the extrusion process along the gland leads to the formation of a well-organized, insoluble, silk filament (Silva-Zacarin et al. 2003). However, the primary and secondary structure of honey bee silk is very different to that of most other silks. Honey bee silk comprises comparatively low levels of glycine residues whereas levels of alanine, serine, aspartic and glutamic acids are elevated (Lucas and Rudall 1968). While extensive β-sheet strands make up the main part of silkworm silk and spider silks, honey bee silk shows mainly α-helical character. The helices are assembled into a coiled coil structure running parallel to the fiber axis. One further study indicated the presence of tetrameric coiled coils (Atkins 1967; Fig. 16.4). Coiled coils are a widely distributed structural motif in which multiple α-helices with a characteristic seven-residue periodicity wind around each other to prevent hydration of hydrophobic residues buried in the core of the structure (Woolfson 2005). These heptad repeat sequences
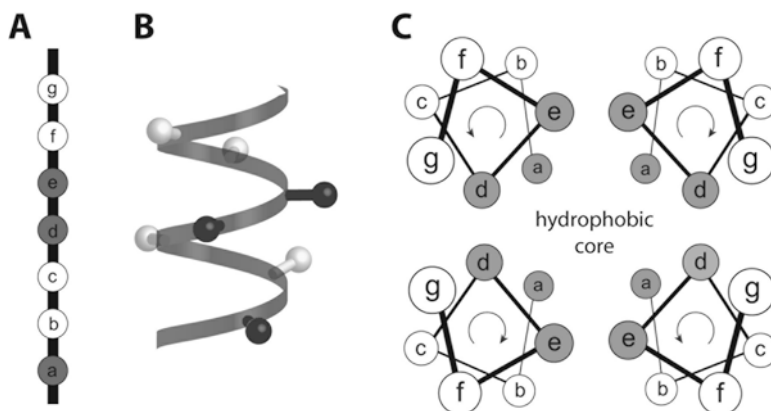
Fig. 16.4 The tetrameric coiled coil structure of honeybee silk. (**a**) The primary structure of honeybee silk consists of heptad repeats, where the positions *a*, *d* and *e* (*grey*) are occupied by predominantly hydrophobic amino acid residues. (**b**) Arranged into α-helical structure, the residues *a*, *d* and *e* point into similar directions. (**c**) Assembled into antiparallel tetrameric configuration, the hydrophobic residues are buried in the core

are generally denoted as $(abcdefg)_n$. Large hydrophobic residues occupy positions *a* and *d*, whereas the more polar and charged residues reside in the other positions (Woolfson 2005).

Comparison of several coiled coil fibroin proteins revealed a low degree of sequence identity among different species and no extensive repeats, while, in contrast, the repetitive primary sequence of silk proteins creating β-sheet silk is rather conserved. The sequence diversity might be a result of repeated amino acid substitutions with amino acids possessing similar properties during evolution. The exact sequence seems less important than the conservation of the character of the side chains in special heptad positions (Sutherland et al. 2007). As the four identified silk genes (see above) are expressed at similar rates in the larvae, it is highly probable that all four proteins are involved in the four strands of the predicted coiled coil structure (Weisman et al. 2010). Interestingly, relatively large amounts of alanine are found in the honey bee silk *a* and *d* position although alanine is a less favored residue in this position. Alanine alone is classified as not hydrophobic enough to stabilize the coiled coil conformation, indicating the influence of further factors on secondary structure formation (Sutherland et al. 2006). Interestingly, no cysteines are present in honey bee silks. The relatively high abundance of alanine among polar and charged residues in the remaining positions might be interpreted as an adaption of honey bee silk to the waxen environment in the hive. While many other coiled coil structures are found in an aqueous environment, honey bee silk is used in a highly hydrophobic one (Sutherland et al. 2006).

16.2.1.2 Recombinant Honey Bee Silk Production

Honey bee silk proteins are suitable for recombinant expression in *E. coli* as these proteins generally do not comprise highly repetitive sequences and they have a small molecular weight of ~30–45 kDa (Sutherland et al. 2012; Weisman et al. 2010). The proteins consist of extensive coiled coil regions flanked by short termini. The hydrophilic termini might increase the solubility of the silk proteins in the gland prior to spinning and might therefore have a similar function as the terminal regions of β-sheet silks (Sutherland et al. 2007), which will be discussed below in detail.

The silk proteins of the honey bee *Apis mellifera* are encoded by four small (~ 1 kb each) and non-repetitive genes (Weisman et al. 2010; Sutherland et al. 2006). For recombinant protein production, the four genes have been cloned from the cDNA clones into pET expression vectors without their signal peptides, and *E. coli* Rosetta 2 (DE3) served as expression host. The bacteria were cultured to an optical density at 600 nm (OD_{600}) of ~20, and gene expression was induced for further ~24 hours. As an example, AmelF3 could be purified from inclusion bodies at a yield of 2.5 g per liter. Also AmelF1, 2 and 4 could be produced recombinantly and purified at yields of 0.2, 1.5 and 1.9 g per liter, respectively (Weisman et al. 2010). Studies revealed that solutions of one protein alone were less stable than solutions of equimolar mixtures of all four proteins. Fibers hand-drawn from the mixture were comparatively extensible but had lower tensile strength in comparison to natural honey bee fibers. The fibers were not birefringent suggesting that the proteins were not aligned. Post-stretching of the fibers in methanol leads to birefringence similar to native fibers indicating the alignment of the proteins within the fiber, and, though extensibility was reduced, the tensile strength of native fibers could be mimicked (Weisman et al. 2010).

Due to the knowledge of the *A. melliferis* silk genes, homologous genes could be identified in the Asiatic honey bee *Apis cerana (*Shi et al. 2008*)*. With 289 to 321 amino acid residues, these proteins exhibit a similar size compared to that of the European honey bee. For recombinant protein production, the respective four DNA fragments encoding the Asiatic honey bee silk proteins 1–4 (ABS1–4) were used as a template for amplification by PCR. The primers were designed to amplify the genetic sequence of the silk proteins without their signal peptides. Resulting PCR products were cloned into pET expression vectors providing His-tags to the proteins. *E. coli* strain BL21 (DE3) served as an expression host. Bacteria were cultivated to an OD_{600} of 0.7, and gene expression was induced for three hours. The soluble protein was purified by affinity chromatography at yields of 30, 30, 10 and 60 mg/ml, respectively. The work of this group could in the first instance not confirm the expected coiled coil structure. In contrast, in aqueous solution, mainly a random coil conformation was reported (Shi et al. 2008). Further research has to reveal if this is an effect of sequence changes between European and Asiatic honey bee silk or if it might be attributed to the influence of the His-tag or processing conditions.

The recombinant production of four silk proteins of the giant honey bee *Apis dorsata* further shed light on the folding behavior of the silk proteins (Maitip et al. 2015). Recombinantly produced, His-tagged AdorF1-4 silk proteins were found to be soluble without the addition of SDS or other detergents. This allowed more precise investigations of the folding behavior of these silk proteins. The proteins were of similar size and amino acid composition as the recombinant proteins mentioned above. Interestingly, the folding behavior was different for equimolar mixtures of all four proteins compared to individual silk solutions. The CD (circular dichroism) spectra revealed that proteins were structured as coiled coils above concentrations of 3 mg/ml. At lower concentrations, solutions of individual proteins started to dissociate into unassociated α-helices, whereas the mixed solution retained the coiled coil structure (Maitip et al. 2015). Tetrameric coiled coils are therefore more stable than homomeric coiled coils, indicating that the four proteins probably are not functionally redundant, but each have a critical role.

16.2.2 Examples of Biotechnologically Produced Silk Proteins: Lacewing Silk

16.2.2.1 Natural Lacewing Silk

The coiled coil motif in silk has evolved convergently in at least five insect lineages - the stinging hymenopterans (ants, bees, hornets), sawflies, fleas, praying mantises and lacewings (Sutherland et al. 2014; Kameda et al. 2014). Larvae of the green lacewing (*Mallada signata*) produce a cocoon silk protein in their Malpighian tubules that is 49 kDa in size and rich in alanines (Weisman et al. 2008). The final instar lacewing larvae of both sexes spin a loosely woven silk cocoon to pupate inside (Weisman et al. 2008). The outer silk layers attach the construct to a substrate and also provide a sticky surface to attach debris or leaves to it in order to disguise the cocoon from predators, while the inner layer serves as a scaffold for lipid coverage (Kameda et al. 2014). Although experimental data on the cocoon silk's secondary structure are poor and rather indicate unassociated α-helical structures, bioinformatics analysis provided evidence for the presence of heptad repeats supporting the coiled coil structure, although the close proximity of lipids in the composite material could lead to a destabilization of coiled coils in the hydrophobic environment (Weisman et al. 2008).

Interestingly, in the adult live stage, lacewings produce a completely different, unrelated silk. Although adult insects in general hardly produce silk, female lacewings produce egg stalk silk from collateral glands (Rudall and Kenchington 1971). In order to protect the eggs from predation, lacewings place their eggs underneath leaves or other kind of shelter. The female secrets a drop of silk dope onto a surface (e.g. of a leaf) and dips in an egg. By pulling the abdomen away, it creates the thin egg stalk on top of which the egg resides. The egg stalk hardens within few seconds, and the egg is released completely. Already in 1957 a structural peculiarity had been

published. In contrast to spider silk or silkworm silk which consists in parts of parallel β-sheet structures, cross-β-structures with individual β-strands aligned perpendicular to the fiber axis were reported in lacewing egg stalks (Parker and Rudall 1957). This secondary structure arrangement leads to comparably high bending stiffness of the stalks and makes them particularly resistant to transversal forces. Egg stalks of *M. signata* exhibit a bending modulus nearly three times higher than that of silkworm fibers (Weisman et al. 2009). The extensibility of egg stalks is highly humidity-dependent: while an extensibility of up to 600 % is reported at high humidity, at a relative humidity of 30 % only values of around 5 % are achieved (Weisman et al. 2009; Bauer and Scheibel 2012).

Based on the secondary structure of lacewing silk, a model has been proposed to explain the humidity-dependent performance of the material (Bauer et al. 2012). The effects were mainly ascribed to the transition from cross-β-structures to parallel β-sheets based on the interplay of disulfide bridges crosslinking protein molecules of different β-strands and the variable strength of stabilizing hydrogen bonds in the presence or absence of water. At low humidity, the disulfide bonds break up already at low strain as the strength of the hydrogen bonds in total is higher than the strength of the disulfides, and the stalk can take up only little energy. At high humidity, residual water molecules function as new hydrogen bond donors/acceptors competing with the hydrogen bonds between β-sheets and weakening the cohesion of the β-sheets via hydrogen bonds. In this situation, the disulfides are comparatively strong, and, upon extension of the stalk, breakage of hydrogen bonds between the β-sheets allows an irreversible rearrangement of the cross-β-structure into parallel β-sheets, followed by a significant increase of extensibility of the stalk and thus the ability to absorb much more energy (Bauer et al. 2012).

The fascinating properties of natural lacewing egg stalks encouraged further research (albeit not promoted at the pace of research on silkworm silk or spider silk) on the underlying proteins (Lintz and Scheibel 2013). The silk dope of *Chrysopa carnea* has been shown to consist of five proteins. In the species *Mallada signata* two egg stalk proteins, namely MalXB1 (86 kDa) and MalXB2 (55.5 kDa) have been identified (Weisman et al. 2009). These serine- and glycine-rich proteins consist of 51 or 29 repeats of a 16 amino acid long motif, respectively, which are flanked by non-repetitive terminal domains. Additionally, MalXB1 has a non-repetitive region in the middle of the protein. MalXB1 contains seven, MalXB2 five cysteine residues found predominantly in the terminal domains (Weisman et al. 2009).

16.2.2.2 Recombinant Lacewing Silk

A 48 amino acid long consensus motif of the core domain of MalXB2, named [AS], was selected to be recombinantly produced in *E. coli* (Bauer and Scheibel 2012) (Fig. 16.5). The DNA encoding this motif was multimerized eight times and fused to the genetic sequence encoding the natural amino (N) and carboxy (C) terminal domains. The recombinant 53 kDa protein N[AS]$_8$C was purified from *E. coli* and

Fig. 16.5 Recombinant lacewing eggstalk silk protein N[AS]8C. Amino and carboxy terminal domains (denoted as N and C) flank eight repeats of the consensus motif AS which has been derived from the core domain of the natural eggstalk protein MalXB2

analyzed regarding its assembly properties. Post-treated N[AS]$_8$C fibers showed very similar secondary structures compared to natural egg stalks as measured by FTIR spectroscopy.

16.2.3 Examples of Biotechnologically Produced Silk Proteins: Caddisfly Silk

16.2.3.1 Natural Caddisfly Silk

Caddisflies (order Trichoptera) are a large group of (partially) aquatic insects with over 12,000 species (Wiggins 2004). The larvae spend their lives in freshwater streams and lakes where they feed, mature and pupate under water (Bai et al. 2015; Ashton et al. 2013). After several larval stages the larvae undergo complete metamorphosis inside their underwater cocoon before hatching as a short-lived, winged adult to leave the water for terrestrial mating (Ashton et al. 2013; Stewart and Wang 2010).

Caddisflies are closely related to Lepidoptera, an order including silk-spinning terrestrial moths and butterflies (Ashton et al. 2013; Morse 1997). They diverged from a common ancestor and successfully penetrated diverse aquatic habitats largely owing to the development of adhesive underwater silk fibers (Ashton et al. 2013; Stewart and Wang 2010). In order to provide shelter for their soft bodies, casemaker larvae use their silk like an adhesive tape to gather organic material, e.g. debris, small stones or sand grains, twigs and leaves to build a protective case, while retreat builders create stationary composite retreats attached to firm ground in streaming water which can additionally be equipped with silk nets to catch prey (Ashton et al. 2013; Bai et al. 2015).

Regarding the exceptional ability of this silk to stick to any surface underwater, it could serve as a natural blueprint for biocompatible water-borne adhesives (Stewart 2011; Brubaker and Messersmith 2012). Obtaining this sticky silk on an industrial scale would be highly interesting for biomedical as well as technical applications (Brubaker and Messersmith 2012). Efficient adhesion under wet conditions is required in many fields, including surgery when repairing living wet tissue (e.g. gluing tissues or bones together) or in dentistry, but also for the replacement

for any technical glue which has to be used in wet or aquatic environment. Until now, most commercially available glues fail to perform in wet environments. Therefore, caddisfly silk has attracted researchers' interest as a model for a biopolymeric glue.

Similar to *Bombyx mori* silkworm silk (see below), caddisfly silk is spun in a pair of modified salivary glands as fused fibers (Tsukada et al. 2010). In both orders, the silk is mainly made up from H- and L-fibroins (heavy and light chain fibroins) with a molecular weight of more than 300 kDa and roughly 25 kDa, respectively (Ashton et al. 2013). The H-fibroins comprise a long, repetitive core region which is flanked by short non-repetitive termini (Ashton et al. 2013). The highly conserved cysteine residues might be a hint for covalent linkage of the two proteins in a stoichiometric manner (Ashton et al. 2013). Despite these similarities, distinct molecular adaptation of the dry silk was necessary to create underwater adhesives. Comparing the primary structure of terrestrial silks to that of caddisfly silk reveals the accumulation of basic residues, especially arginine, which has been attributed to the adaption of caddisfly silk to aquatic systems (Stewart and Wang 2010). Further characterization showed that these silks lack the abundant poly(A) and poly(GAGAGS) domains seen in silkworm and spider silk, which form rigid β-sheet structures and contribute to the mechanical properties (Marsh et al. 1955; Holland et al. 2008). Instead, extensive $(SX)_n$ repeats were found in the sequence of H-fibroin with S indicating serine, X an aliphatic amino acid, often valine or isoleucine (which both are known to play a role in β-sheet formation (Johansson et al. 2010)) and n numbers from 2 to 6 (Yonemura et al. 2006; Wang et al. 2010; Ashton et al. 2013). These $(SX)_n$ stretches are hypothesized to account for the β-sheet structure of caddisfly silk. Moreover, in proximity to these $(SX)_n$ structures, conserved proline-glycine (PG) motifs are found which can also be seen in β-turns (Jenkins et al. 2010). This is in accordance with the observation that X-ray diffraction studies revealed a three-sheet ordered structure in the absence of poly(A) stretches (Engster 1976). It has been estimated that more than half of the serines are phosphorylated (Stewart and Wang 2010). The phosphorylations are mainly found in the prevalent $(SX)_4$ repeats. Together with the relatively high levels of multivalent cations such as Ca^{2+}, Mg^{2+} and Fe^{3+} detected in caddisfly silk, a model was developed with alternate phosphorylated serine residues interacting with these divalent cations to form the observed β-sheet structure (Stewart and Wang 2010; Addison et al. 2014; Fig. 16.6).

It is assumed that dehydration of the predominantly hydrophilic silk is achieved by crosslinking of phosphoserine-rich domains with metal-ions (predominantly Ca^{2+}) when silk is submerged in water, since the solubility of these ions is low at neutral pH (Addison et al. 2013). This hypothesis is confirmed by a decrease in β-structures upon depletion of Ca^{2+}-ions with EDTA (Ashton et al. 2013). Stress-strain-profiles confirmed a loss of toughness and a dramatically reduced young's modulus upon the addition of EDTA to natural silk fibers. An additional adaptation of caddisfly silk can be seen in the substantial hysteresis effect. Controlled cylindrical strains were applied to caddisfly silk fibers, and their stress response revealed normal loading profiles. But interestingly, the unloading profile showed nearly full recovery of the initial modulus and yield point, although the strain was beyond the
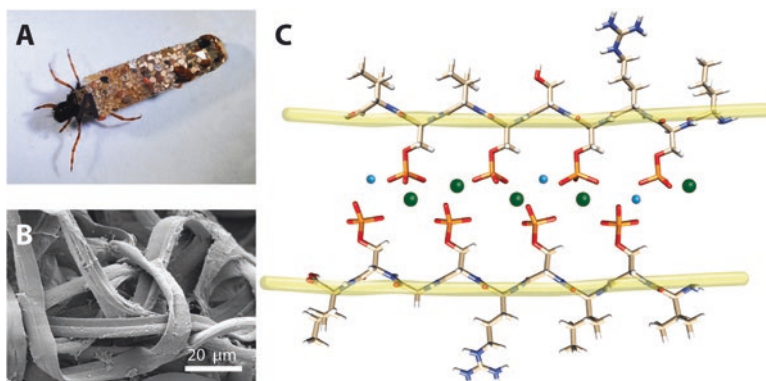
Fig. 16.6 Caddisfly silk. (**a**) Caddisfly larvae with a silk case and attached debris. (**b**) SEM micrograph of caddisfly silk. (**c**) Model for the structural motif of phosphorylated $(SX)_4$ repeat regions, in which divalent cations **Ca^{2+}** and **Mg^{2+}** (*green and blue dots*) complex with negatively charged phosphoserine residues (Reprinted and adapted with permission from (Addison et al. 2013). Copyright (2013) American Chemical Society)

(pseudo)yield point. This fact indicated that the deformation was reversible, while the strain energy dissipated as heat (Ashton et al. 2013). Such a self-tensioning effect might be very helpful for the larvae to keep stones and twigs tightly assembled on their body. The viscoelasticity of caddisfly silk has been attributed to two independent Ca^{2+} networks: in addition to Ca^{2+}-stabilized phosphoserines of H fibroins, Ca^{2+}-complexed carboxylate groups were found in the amino terminal region (Ashton and Stewart 2015).

16.2.3.2 Recombinant Caddisfly Silk

Biotechnologically, it is difficult to produce caddisfly silk proteins. More generally, the recombinant production of protein-based adhesives has shown only limited success (Stewart 2011). For efficient adherence to underwater substrates, many aquatic organisms need to extensively post-translationally modify their amino acid residues. For example, mussel foot proteins 3 and 5 (mfp-3 and -5) contain more than 20 % of hydroxylated tyrosines (Papov et al. 1995). For caddisfly silk, massive phosphorylation of serines has been shown (Stewart and Wang 2010). Generally, polymeric phosphates are well-known adhesion promotors and are already applied in dental treatments and in the coating industry (Stewart and Wang 2010). If these bio-adhesive proteins are produced by bacterial hosts such as *E. coli*, no substantial post-translational modifications (PTMs) occur.

In order to circumvent the obstacles of heterologous protein production, some effort has been made on the approach to engineer proteins and polymers to imitate properties of underwater adhesives and transfer these properties of the natural system to polymers (Stewart 2011). Just recently, the viscoelastic, tough and

self-strengthening properties of caddisfly silk have served as a model to toughen synthetic polymeric hydrogels. To mimic its H-fibroin, a phosphate-graft-methacrylate pre-polymer was copolymerized with polyacrylamide. If a critical phosphate side chain density was reached, equilibration with Ca^{2+} led to an increased initial stiffness, and a 100-fold increased work-to-fracture was needed compared to hydrogels equilibrated with other cations such as Na^+ or Mg^{2+}. This finding was generally assigned to the energy-dissipating, reversible unfolding of phosphate-metal ion crosslinks (Lane et al. 2015). Caddisfly silk-inspired toughening of hydrogels provides a simple model for further structure-function investigations and perhaps allows the development of synthetic soft materials which exceed the mechanical properties of natural materials (Lane et al. 2015; Ashton and Stewart 2015).

Just recently four major *S. marmorata* silk proteins (Smsps 1–4) have successfully been identified and recombinantly produced in *E. coli*. Analysis of the amino acid sequence of these proteins revealed the presence of a new GYD-rich motif which could play an important role for the mechanical properties of caddisfly silk (Bai et al. 2015), since di-tyrosine crosslinking has already been reported for the underwater adhesive silk of *H. occidentalis (*Wang et al. 2014*)*.

16.2.4 Examples of Biotechnologically Produced Silk Proteins: Silkworm Silk

16.2.4.1 Natural Silkworm Silk

Silk from silkworm has been used since ancient times and is still used today (e.g. to manufacture cloth). Silkworms, the larvae of the holometabolous silk moths, undergo several molts during their life cycle before they pupate inside their silken cocoon. Upon completion of metamorphosis, the release of proteolytic enzymes leads to the partial destruction of the cocoon and allows the adult insect to escape. Originally, silk was harvested from collected cocoons of wild silk moths, but the silk of these cocoons generally varied in color and structure, and availability was low. Upon breeding, the domesticated silkmoth *Bombyx mori* emerged; the larvae of this moth produce large amounts of silk within their silk glands (Tomita 2011; Xu 2014).

Silk from the silkworm *B. mori* has not only been used for textile production but has also been used as a biomaterial. Owing to its extraordinary mechanical properties and biocompatibility, it has been used in biomedical applications, e.g. as suture material, for centuries (Altman et al. 2003). Even though some biocompatibility problems have been reported, this finding is likely attributable to the incomplete or omitted removal of sericin, as affirmed by studies on silkworm fibroins indicating a biocompatibility of silkworm fibroin comparable to that of other biomaterials, such as polylactic acid or collagen (Altman et al. 2003). However, biocompatibility alone is not sufficient for successful application of silkworm silk in e.g. tissue engineering.

Silkworms produce their silk from a pair of labial glands from where the silk emerges as a double thread. The silk is mainly composed of fibroin coated with sericin, a family of glue-like glycoproteins. Silk fibroin consists of heavy (H)-chain fibroin, a hydrophobic protein with a molecular weight of 350 kDa which forms a complex with the relatively hydrophilic 25 kDa fibroin light (L)-chain (Yamaguchi et al. 1989) via a disulfide bridge (Tanaka et al. 1999b). P25, a 30 kDa glycoprotein (Tanaka et al. 1999a) associates non-covalently with the disulfide-linked fibroin complex mainly driven by hydrophobic interactions with the H-chain (Tanaka et al. 1999a, 1993). Quantitatively, the ratio of heavy chain, light chain and P25 is 6:6:1 (Inoue et al. 2000). In the caterpillar, the fibroins are mainly secreted into the lumen of the posterior silk gland from where they are transported to the middle silk gland to add sericin and further towards the anterior silk gland where the silk is spun into fibers (Inoue et al. 2000).

16.2.4.2   Recombinant Silkworm Silk

Despite the natural availability of silkworm silk, *B. mori* silk fibroins have also been produced recombinantly in expression hosts, such as *E. coli (Ohshima and Suzuki*1977*)*. Full-length as well as partial silk genes were transformed into *E. coli* and analyzed. Due to the size of the silk proteins massive deletions were found in the repetitive domain. Later on, synthetic proteins based on specific motifs of silkworm silk were produced in *E. coli*. Even hybrid proteins were designed where a sequence derived from the crystalline domain $(GAGAGS)_n$ of *Bombyx mori* silk fibroin was fused to the cell-adhesive sequence TGRGDSPA originating from fibronectin and additionally fused to the sequence $(GVPGV)_n$ from elastin to tailor properties of fibers spun therefrom (Yang et al. 2008).

Interestingly, good expression of modified silkworm genes was shown using silkworms as expression systems. Silkworms have been used to produce foreign proteins already in 1985, when a *B. mori* nucleopolyhedrovirus (BmNPV) expression system enabled the heterologous production of human α-interferon in silkworm larvae (Maeda et al. 1985). However, this transient expression system is limited to one generation, and, therefore, the infection with this particular virus has to be carried out repeatedly.

Great progress in recombinant protein production in silkworm silk glands has been made upon the development of a stable germline transformation system for *B. mori*. A vector derived from a lepidopteran transposon, *piggyBac*, was shown to be able to transpose genetic information into the silkworm's chromosome, thereby generating stable germline transformations (Tamura et al. 2000). Genetic sequences encoding peptides from collagen or fibronectin were fused to the *B. mori* L-chain fibroin gene and introduced into *B. mori* using the *piggyBac* transposon (Yanagisawa et al. 2007). The first peptide, called Coll-F, consisted of a consensus sequence derived from collagen $[GERGDLGPQGIAGQRGVV(GER)_3GAS]_8$ fused to the crosslinking sequence GPPGPCCGGG derived from collagen III to promote triple helix formation through disulfide bonds formed between the chains (Yao et al.

2004). Another peptide, abbreviated RGD, with the sequence [TGRGDSPAS]$_8$ was derived from fibronectin. The transformation vectors pigFiLpColl-FGFP/3xP3DsRed for the production of fibroin containing the genetic information for L-chain, Coll-F and GFP was constructed as well as the construct containing the RGD sequence. For both constructs, a transgenic silkworm line could be established, and the recombinant silk could be analyzed concerning cell adhesion of mouse fibroblasts. Surprisingly, the recombinant proteins were not only found in the fibroin but also in the sericin layer.

Another example of engineered silkworm silk is the production of chimeric silkworm/spider silk fibers for improved mechanical properties. A spider silk derived protein called A2S8$_{14}$ comprised a (GPGGA)$_8$ sequence and a poly-alanine sequence (linker-alanine$_8$). The encoding gene was cloned into a piggyBac vector further containing the *B. mori* fibroin heavy chain promoter to target protein secretion into the posterior silk gland, as this promotor is active in this part of the gland (Sehnal and Akai 1990). The sequence encoding the synthetic spider silk protein (78 kDa or 106 kDa when fused to GFP) was cloned between the amino and carboxy terminal domains of the *B. mori* fhc protein and finally, transgenic silkworms could be established and their silk analyzed with regard to the mechanical properties.

16.2.5 *Examples of Biotechnologically Produced Silk Proteins: Spider Silk*

16.2.5.1 **Natural Spider Silk**

Arthropods have evolved a huge variety of different silks. Besides the insect silks mentioned above, spiders (Araneae) are prominent producers of silks. Spiders produce silks for a great variety of purposes, such as to catch prey, to wrap it afterwards, to protect their offspring or as a lifeline. Interestingly, not all spiders produce spider silk webs to catch prey; among all known spider species only some do so (Römer and Scheibel 2008). Evolution has brought a plethora of different web shapes, with orb webs being the best studied ones (Römer and Scheibel 2008). Female orb weaving spiders can produce up to seven different silk types from separate glands, all adapted and evolutionarily optimized to fulfill their respective tasks in the web (Vollrath 2000; Fig. 16.7). The outer frame and the radii of an orb web are constructed from a silk with high tensile strength (Blackledge et al. 2005; Gosline et al. 1984; Vollrath and Porter 2006). This silk is mainly made of two types of silk proteins originating from the major ampullate gland, therefore the fiber is called major ampullate silk. This silk also functions as an always-ready lifeline which is dragged behind to facilitate rapid escape from predators and therefore has the by-name "dragline silk". In contrast, the capture spiral of the web is made of highly elastic flagelliform silk (produced in the flagelliform gland) to absorb and dissipate the enormous kinetic impact on the web if prey is caught in mid-flight (Becker et al. 2003; Dicko et al. 2004). Silk proteins produced from the minor
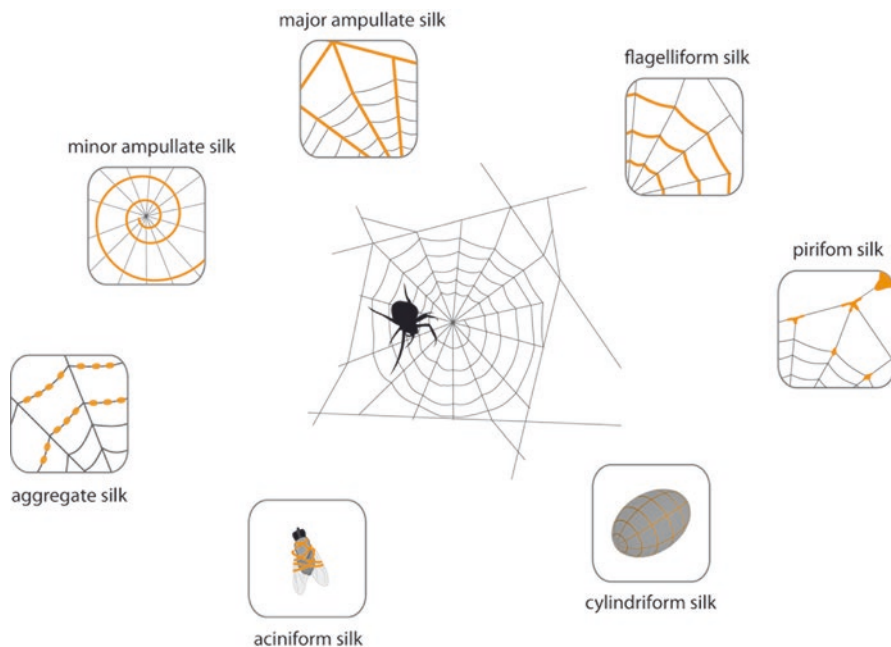
Fig. 16.7 Different types of spider silks spun by female orb weaving spiders

ampullate silk gland are used in auxiliary fibers during construction of the web (Dicko et al. 2004; Riekel and Vollrath 2001). Proteins from the piriform silk gland provide a sophisticated cement to attach both, fibers to each other and the web to a substrate (Hajer and Rehakova 2003). Aciniform silk and cylindrical silk protect the eggs (Garb and Hayashi 2005; Hayashi et al. 2004). And, finally, aggregate silk serves as a sticky coating on the capture spiral (Vollrath 2006; Hu et al. 2007).

Among the orb weaving spider silks, draglines of *Nephila clavipes* and *Araneus diadematus* are the best studied ones (Scheibel 2004). On a weight-to-weight basis, the tensile strength of some silk fibers is comparable to that of steel, and the elasticity of others comparable to rubber (Römer and Scheibel 2008). Remarkably, spider dragline silk is classified as the toughest natural fiber and thus superior to other biological materials, such as silkworm silk, and even superior to man-made high-performance fibers, such as Nylon or the aramid-based fiber Kevlar (Liu et al. 2005; Omenetto and Kaplan 2010; Porter et al. 2013; Tokareva et al. 2013). Especially, the combination of elasticity and strength and low weight of the fiber at the same time has attracted researcher's interest and prompted extensive investigation on structure-function relationships, as the mechanical properties of different silks can be ascribed to the underlying amino acid sequence.

Generally, spider silk proteins comprise three domains: a large, highly repetitive core domain flanked by non-repetitive amino and carboxy terminal domains (Huemmerich et al. 2004a). The repetitive sequences are composed of up to over 100 repeats of short polypeptide stretches of a length of up to 60 amino acids and

account for up to 90 % of the protein (Huemmerich et al. 2004a; Guerette et al. 1996). Spider silk proteins are large proteins with a size in the range of 250–320 kDa (Ayoub et al. 2007; Sponner et al. 2005). Typical motifs seen in dragline silk are blocks of alanines, which assemble into crystalline β-sheet stacks and account for the strength of the fiber (Hijirida et al. 1996; Parkhe et al. 1997), whereas glycine-rich motifs such as GGX or GPGXX confer elasticity to the fiber through adopting 3_{10}-helical structures or β-turn spirals, respectively (Van Beek et al. 2002; Hijirida et al. 1996). The primary structure of spider silk protein core domains characteristically comprises alternating patterns of hydrophilic and hydrophobic blocks (Römer and Scheibel 2008). While the core domain determines the macromolecular properties of the resulting protein fiber and, therefore, is more variable between silk types and spider species, the terminal domains are highly conserved and consist of 100–200 amino acids with α-helical structures in a five helix bundle (Rising et al. 2006; Challis et al. 2006). The terminal domains play a key role in storage and fiber assembly (Askarieh et al. 2010; Hagn et al. 2010, 2011).

16.2.5.2 Recombinant Spider Silk

Spider silks are interesting candidates for many technical as well as biomedical applications. In contrast to silkworms like *Bombyx mori*, farming of spiders as a living source of silk is impeded by their aggressive territorial and cannibalistic nature. In addition, silk from spiders kept in captivity fluctuates in quality, i.e. the nutrition as well as the general condition of the animal has significant influence thereon (Madsen et al. 1999). Recombinant production of spider silk proteins allows the production of large amounts of protein with constant quality, and additionally facilitates genetic modifications of the proteins.

First attempts to express native spider silk genes in *E. coli* have had only limited success (Arcidiacono et al. 1998). The repetitive sequences and high G C content made manipulation of the genes difficult, and exemplarily silk genes of *Nephila clavipes* of a size larger than 2.5 kb exhibited genetic instability and underwent recombination (Arcidiacono et al. 1998). Therefore, synthetic silk-like genes were developed (Prince et al. 1995; Lewis et al. 1996). One of the first approaches to produce spider silk proteins in *E. coli* was reported in 1995 (Prince et al. 1995). Consensus motifs were derived from the available cDNA sequences of *N. clavipes* and back-translated into the respective DNA sequences regarding the respective codon usage of the desired expression host *E. coli*. Exemplarily, the consensus sequence GGAGQGGYGGLGSQGAGRGGLGGQGAG was multimerized with a sequence of ten alanine residues in between each block to mimic the character of spider silk proteins. As, at that time, synthesis of entire genes was not far advanced, the sequence coding for the consensus motif had to be assembled from four separate oligonucleotides by utilizing 5′-*Nhe*I and 3′-*Spe*I ends. Finally, the sequences were sub-cloned from the cloning vector pUC18 into pQE-9 or pET-21a expression vectors, and the gene expression was induced in *E. coli* bacteria transformed with this vector upon the addition of IPTG. Multimers of 4, 7, 10, and 13 repeats were

successfully purified, but the protein yield after purification was as little as 2–15 mg per liter. Later on a plethora of partial spidroins as well as spider silk-derived proteins were successfully produced in *E. coli* (Heidebrecht and Scheibel 2013; Tokareva et al. 2013). Limitations such as low yields, truncated proteins due to premature translation termination e.g. as a result of tRNA depletion due to G and C rich genes or the accumulation of silk proteins in inclusion bodies had to be managed (Vendrely and Scheibel 2007). A large proportion of recombinant proteins have been in the range of 10–100 kDa, which seems to give good results concerning protein quality and yield. However, the size of a silk protein is suspected to have an influence on the mechanical properties of silk fibers spun therefrom. In 2010, the recombinant production of a native-sized *Nephila clavipes* silk protein was reported in metabolically engineered *E. coli* (Xia et al. 2010). The partial cDNA sequence of Major ampullate spidroin I (MaSp1) of *N. clavipes* was the basis for the consensus motif SGRGGLGGQGAGMAAAAAMGGAGQGGYGGL GSQGT. Mulitmerization of the module by the "head-to-tail" strategy created constructs with 32–96 repeats of the module resulting in a predicted molecular weight from 100.7 to 284.9 kDa (including a His-tag and S-tag at the amino terminus). In a preliminary experiment, an upregulation of a glycine biosynthetic pathway was seen as a stress response in *E. coli* bacteria upon the production of recombinant silk proteins. In order to eliminate this putative bottleneck of insufficient provision of glycine during translation, the bacteria were equipped with a compatible plasmid allowing the expression of tRNAs recognizing two glycine codons GGU and GGC. To additionally increase the protein production, the bacteria were engineered to overexpress serine hydroxymethyltransferase. Among the investigated proteins, these engineering strategies especially increased the protein production of the larger sized proteins, i.e. a 64mer, 80mer and 96mer, and enabled a 10 to 35-fold increased protein production. The bacteria were cultivated under high cell density conditions, and protein yields were calculated to be 1.2 g per liter (Xia et al. 2010).

The limitations of recombinant spider silk protein production seen in *E. coli* have prompted research on expression hosts other than *E. coli* to attain improvements concerning yield, quality and availability. Other bacterial hosts such as *Salmonella typhimurium* have been exploited to produce silk proteins due to their ability to secrete proteins to the extracellular environment (Lee et al. 2006; Widmaier et al. 2009). Therefore, a type III secretion system (T3SS) encoded within the *Salmonella* Pathogenicity Island 1 has been utilized and genetically manipulated to secrete monomers of *Araneus diadematus* spidroins ADF-1, -2 and-3 (Widmaier et al. 2009).

Yeasts, such as the *Pichia pastoris*, are auspicious alternatives for the production of spider silk proteins as they are able to synthesize and secrete large proteins (Cereghino et al. 2002; Cregg et al. 1993). Synthetic spider silk genes were expressed under control of the methanol-inducible AOX1 promoter at high levels, but while proteins encoded by genes with up to 3000 codons were produced successfully, increasing size led to significantly decreased yields, and the size of the proteins varied from the expected ones (Fahnestock and Bedzyk 1997).

Insect cells have been tested for recombinant spider silk production as insects and spiders are from the same phylum (Arthropoda), and, therefore, insect cells are very close phylogenetic relatives compared to other frequently used expression hosts (Huemmerich et al. 2004a). *Araneus diadematus* spidroins ADF3 and ADF4 have been successfully produced in a fall armyworm *Spodoptera frugiperda* derived cell line, Sf9, infected with baculoviruses containing the respective partial cDNA sequences (Huemmerich et al. 2004b). While ADF4 assembled into fibrils in the cytoplasm, ADF3 remained in the soluble state, which was mainly attributed to the differences in hydropathicity. Further, a *N. clavipes* flagelliform silk protein was recombinantly produced in a *B. mori* cell line using a *Bombyx mori* nuclear polyhedrosis virus baculovirus expression system (Miao et al. 2005). The reported yield of the purified 37 kDa protein was 0.08 mg per 6 ml of cell suspension.

Spider silk proteins have also been produced in mammalian cells. These cells are able to produce large proteins and can also secrete the produced protein. Two cell lines, namely bovine mammary epithelial alveolar cells (Kamenskiy et al. 2014) and baby hamster kidney (BHK) cells, have been used for the production of partial dragline silk proteins ADF3 of *Araneus diadematus* and MaSpI and MaSpII of *Nephila clavipes*. Analysis of the medium of stable transfectants of all constructs revealed the successful production and secretion of silk proteins by these cell lines. Nevertheless, multimerization of the ~60 kDa constructs leading to larger protein sizes up to 140 Da was accompanied by much lower yields (Lazaris et al. 2002).

A genetic sequence encoding a part of the major ampullate spidroin 1 from the African spider *Euprosthenops sp.* was cloned into an expression vector suitable for transient transfection of mammalian COS-1 cells, a cell-line derived from monkey kidney (Grip et al. 2006). The constructs contained a carboxy terminal poly-histidine tag to allow affinity purification, and some transformants contained an additional amino terminal leader sequence for protein secretion. Although, successful transient production of the 22.5 kDa and 24.9 kDa constructs could be demonstrated, the expression rate was low, suggesting these cells would encounter similar problems as bacterial expression hosts, such as obstacles with efficient transcription and translation of repetitive sequences (Grip et al. 2006).

Transgenic animals and plants have been created to test them as hosts for recombinant spider silk production. Plants were discovered as promising expression hosts as they offer a plethora of expression loci, e.g. roots, leaves, seeds or fruits (Scheller and Conrad 2005). Transgenic tobacco and potato plants have been investigated concerning the expression of synthetic MaSp1 genes of *Nephila clavipes* of 420 to 3600 bp in size (Scheller et al. 2001). The proteins were produced in the endoplasmatic reticulum of tobacco and potato leaves and potato tubers and were found to represent up to 2 % of the soluble protein, thereby showing that the production of spider silk proteins of up to 100 kDa is feasible. *Arabidopsis thaliana* and somatic soybean embryos served as an expression host of an *N. clavipes* MaSp1-derived gene in the leaves and seeds (Barr et al. 2004). While smaller silk-like proteins were produced efficiently, production of larger proteins of 127 kDa led to the additional production of smaller protein fragments. In another approach, a native sized flagelliform protein of *N. clavipes* has been produced in the endoplasmic reticulum of tobacco leaves (Hauptmann et al. 2013). Exploiting an intein-based posttransla-

tional protein fusion technology, silk protein multimers larger than 250 kDa were formed subsequent to ligation of monomers. This method copes with highly repetitive genes and, therefore, might lead to reduced genetic instability. In further work, intein-based *trans*-splicing led to seed-specific production of synthetic flagelliform silk proteins (Weichert et al. 2016). The seed-specific expression was driven by the unknown seed protein (USP) promoter, and multimers larger than 460 kDa have been reported, which are stable when storing the seeds up to 1 year under ambient conditions.

Also, genetically modified silkworms (*B. mori*) were established using a Bac-to-Bac/BmNPV baculovirus expression system to produce the dragline silk protein MaSp1 of *Nephila clavata* (Zhang et al. 2008). The genetic fragment encoding a 70 kDa protein was fused to an EGFP sequence. In a larva, approximately 6 mg of fusion protein was detected, but 60 % of the recombinant protein was insoluble. In another experiment, a partial cDNA encoding an 83 kDa MaSp1 protein of *Nephila clavata* was cloned into a *piggyBac* vector, and eggs of silkworms were transformed with this vector to generate germline-transgenic silkworms (Wen et al. 2010). The silken pupation cocoons were harvested and analyzed with regard to their design and properties. The silk proteins were located in the sericin layer of the silkworm silk. Mechanical tests revealed a slight improvement of transgenic silkworm silk compared to unmodified silkworm silk, but the mechanical properties did not reach the extraordinary properties of the natural blueprint.

Production of spider silk proteins in transgenic animals has been tested in mammary glands of mice and goats. Based on the partial cDNA of MaSp1 and MaSp2 of *N. clavipes* an artificial gene construct was cloned into the pBC1 vector and microinjected into the pronuclei of fertilized Kunming White eggs. The protein yield from milk was demonstrated to be up to 11.7 mg per liter, but protein size varied (Xu et al. 2007). Moreover, genetically engineered goats were generated to produce recombinant spider silk proteins based on MaSp1 and MaSp2 of *N. clavipes* (Lazaris et al. 2002). Although, spider silk proteins could be produced successfully, these expression systems showed several drawbacks as the generation of transgenic animals is difficult, the generation time is very long and separation of silk proteins from milk caseins is another hurdle for efficient silk protein production (Xu et al. 2007; Heim et al. 2009).

16.3   Processing and Properties of Silk Proteins

In general, there are two possibilities for processing polymeric materials – from the melt or from solution. Silk proteins have exclusively been processed from solution due to their physicochemical properties. The thermal behavior of silk materials has been investigated by differential scanning calorimetry (DSC) and thermogravimetric (TGA) analysis. The results of these observations indicate a bimodal melting and thermal decomposition of silk proteins due to the narrow temperature range in which melting and decomposition occur (Spiess et al. 2011; Vasconcelos et al. 2008; Gotoh et al. 1997). Though it has been shown that, under extreme heating and cooling rates of 2000 K/s, reversible melting and crystallization of silk proteins is
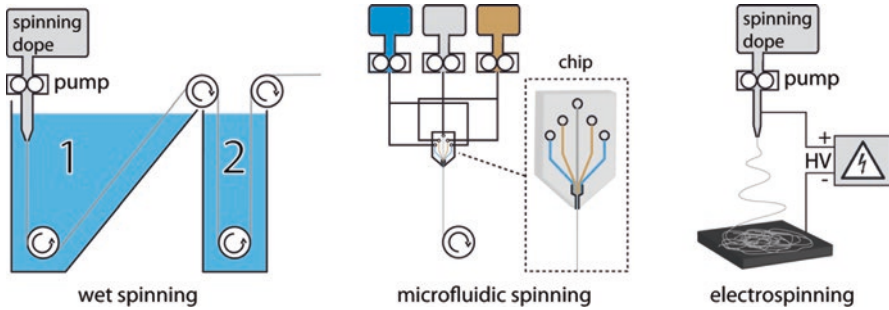
Fig. 16.8 Schematic illustration of different spinning methods, successfully applied to produce artificial silk fibers

possible preventing thermal decomposition (Cebe et al. 2013), it is questionable if such parameters are applicable for established processing procedures. Consequently, the following methods are all solvent based upon dissolution of the proteins in aqueous or organic solvents.

16.3.1 Fibers and Nonwoven Mats

In principle, polymer fibers can be produced by wet spinning, dry spinning, melt spinning, melt blowing, rotational spinning or electrospinning. In the case of silk protein solutions, the most commonly performed methods are wet spinning and electrospinning. However, microfluidic spinning has also been investigated to mimic properties as obtained in natural silk glands (Fig. 16.8).

Wet Spinning The most commonly used method to produce artificial silk fibers with diameters in the micrometer range is wet spinning. Wet spinning is conducted by extruding an aqueous or organic protein solution into a coagulation bath. Solvents which can be used for coagulation baths are polar organic solvents such as methanol, ethanol, acetone, propanol or isopropanol (Chen et al. 2012), as well as aqueous solutions of cosmotropic salts (e.g. potassium phosphate) (Rammensee et al. 2008). In wet spinning, the main objective is the production of fibers with optimal mechanical properties, making wet spinning a highly attractive method for producing silk fibers. Various studies have shown that the mechanical potential of artificial silk fibers can only be approached by mimicking the naturally occurring structure based on high molecular weight proteins, which display small aligned crystals in an amorphous matrix.

   The crucial role of the primary structure is exemplified by the wet spinning of recombinant *N. clavipes* Flag-like silk proteins with compositional variation of the motifs (GPGGX, GGX and a non-iterated spacer) dominating the repetitive core

domain of natural *N. clavipes* Flag proteins (Adrianos et al. 2013). Fibers were wet spun by extrusion of the protein solutions (15 % w/v in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), subsequent addition of 5 % v/v toluene) into pure 2-propanol, and tensile testing was performed. In the presence of the non-iterated spacer, fiber toughness was significantly improved by a factor of ~5.

Additionally, the impact of naturally occurring non-repetitive termini on the mechanics of wet spun spider silk fibers has been examined (Heidebrecht et al. 2015). It was shown that the presence of terminal domains is crucial for silk pre-assembly in the spinning dope. During fiber formation, terminal domains trigger dimerization and are essential to obtain a stable intermolecular protein network. In fact, the effects of this molecular self-assembly of the spidroins enhances toughness even more significantly than increasing the protein's molecular weight (An et al. 2011; Xia et al. 2010; Heidebrecht et al. 2015). However, the most significant factor in the wet-spinning process is post-stretching of the as-spun fibers. Post-stretching is usually performed in alcohol/water mixtures with water acting as a plasticizer allowing the mobilization of glycine-rich regions and therefore promoting conformational switching and chain orientation (Seidel et al. 2000; Lazaris et al. 2002). The significance of post-stretching is exemplified in recent publications (Table 16.2). In all cases, the tensile strength and extensibility are considerably enhanced upon post-stretching, leading to a significant increase of the resulting toughness. Furthermore, it can be seen that the toughest fibers are generated by wet spinning proteins with non-repetitive terminal domains from aqueous solution (Heidebrecht et al. 2015). Unlike solutions with strong denaturants like HFIP, which prevent protein assembly, this approach allows for the biomimetic preassembly of an intermolecular network during wet spinning, which is then further stabilized by molecular alignment during post-stretching.

Microfluidic Spinning Wet spinning approaches have indicated that the performance of artificial silk fibers can be significantly improved by introduction of biomimetic factors (terminal domains, aqueous solutions, induced phase-separation, post-stretching). Hence, another step towards the technical transfer of the natural spinning process is to mimic the geometry of silk glands, as well as the flow profile, and thus to spin fibers with diameters in the range of the natural blueprint. Therefore, a microfluidic device was designed based upon the dimensions of spider silk glands, which allowed for a precise control of chemical and mechanical properties (Rammensee et al. 2008). These outcomes were controlled by altering the elongational flow conditions as well as by introducing two laminar buffer streams, which could be used to induce protein assembly by diffusive ion-exchange and pH-drop. Such a system allowed for a carefully orchestrated interplay of ion-exchange, pH and shear forces. However, a drawback of this system is that it does not allow for dehydration and ion extraction as it is taking place in the natural spinning duct. Microfluidic spinning from aqueous solution (solely triggered by natural assembly mechanisms upon simultaneous fiber formation and stretching) has so far been the closest spinning approach compared with the natural system.

Table 16.2 Mechanical properties of recombinant spider silk protein fibers wet spun into 2-propanol

Silk type	Solvent	c [%w/v]	MW [kDa]	Stretch [%]	d [µm]	σ_{max} [MPa]	ε_{max} [%]	E [MPa]	U_T (MJ/m³)	References
MaSp 1 (N. clavipes)	HFIP	30	46	0	41 ± 3	16 ± 7	1.5 ± 0.4	1060 ± 500	0.11 ± 0.07	AN et al. (2011)
				300	24 ± 5	54 ± 10	4.5 ± 4.0	3110 ± 1400	1.70 ± 2.04	
			70	0	41 ± 4	36 ± 8	3.1 ± 1.8	2780 ± 530	0.87 ± 0.63	
				300	17 ± 5	133 ± 49	23 ± 19	5700 ± 243	23.7 ± 18.5	
Flag-like (N. clavipes)	HFIP[a]	15	66	0	27 ± 1	26 ± 18	0.7 ± 0.4	n.a.	0.06 ± 0.06	Adrinos et al. (2013)
				300	15 ± 1	151 ± 31	85 ± 38	n.a.	89 ± 24	
MaSp 2 (A. aurantia)	HFIP	45–60	87	0	62 ± 3	14 ± 7	1.2 ± 0.5	1880 ± 1910	0.10 ± 0.13	Albertson et al. (2014)
				300	32 ± 8	28 ± 12	93 ± 67	1790 ± 590	26.2 ± 25.8	
MaSp 2 (A. diadematus)	aq[b]	10–17	48	0	35 ± 7	10 ± 2	17 ± 9	900 ± 400	0.9 ± 0.4	Heidebrecht et al. (2015)
				400	15 ± 5	66 ± 22	38 ± 16	2000 ± 900	20 ± 12	
			95	0	70 ± 7	22 ± 4	22 ± 3	400 ± 100	4 ± 1	
				600	22 ± 2	280 ± 48	65 ± 6	4000 ± 600	110 ± 24	
			134[c]	0	155 ± 8	13 ± 2	6 ± 1	500 ± 100	0.3 ± 0.1	
				600	27 ± 10	370 ± 59	110 ± 25	4000 ± 1000	189 ± 33	

[a]subsequent addition of 5 %v/v toluene
[b]dissolved in aqueous solution
[c]dimerization due to additional non-repetitive amino-terminal domain leads to MW of 268 kDa
The data set includes the applied solvent, the protein concentration (c), the molecular weight (MW) of the processed proteins, the extent of post-stretching (stretch), the diameter (d) of the resulting fibers, the tensile strength (σ_{max}) and strain (ε_{max}), the Young's modulus (E) and the toughness (U_T)

    In another approach, a pH-buffered poly(ethylene oxide) solution was used as a shell stream in a microfluidic device to match the viscosity of a regenerated *B. mori*-fibroin solution in water and to mimic the pH-drop in *B. mori* silkworm glands (Kinahan et al. 2011). The device was submerged in a methanol bath, inducing subsequent dehydration and thus further strengthening the silk fibers. Due to the precise parametric control, this system was capable of producing functional silk fibers with tunable properties such as Young's modulus and fiber diameter. However, due to the lack of control over shear stress, post-stretching in alcohol was still necessary to improve mechanical properties and tune fiber diameters. This was also the case in a microfluidic dry spinning setup, where a highly concentrated silk solution (pH 4.8) was pumped through a single channel microchip to create biomimetic shear and elongation conditions (Luo et al. 2014). This data confirmed the dominating influence of post-stretching on the mechanical properties.

Electrospinning Electrospinning is an electrostatically-driven process allowing the production of fibers with diameters ranging from micrometers down to a few nanometers (Sundaray et al. 2004). The basic setup comprises a needle which is charged (0–30 kV) and through which a protein/polymer solution is extruded (Garg and Bowlin 2011). The strong electric field induces electrostatic repulsion at the surface of an extruded droplet and causes Coulombic forces, which ultimately leads to the formation of a Taylor cone. If the electrostatic forces overcome the surface tension, and the molecular cohesion and chain entanglement in the protein/polymer solution is sufficiently high, the droplet is not ejected as a spray, but a jet is formed. At a certain point this jet experiences whipping instabilities, and is thenceforth stretched and dries on its way to an oppositely charged collector plate on which the fibers are deposited as a nonwoven mesh (Reneker and Yarin 2008). The diameters of fibers produced by electrospinning can be well-controlled by adaption of the spinning dope properties (protein/polymer concentration, viscosity, conductivity, volatility and surface tension of the solvent) and process parameters (voltage, distance to counter electrode, flow rate, temperature, humidity) (Greiner and Wendorff 2007). Considering the number of parameters involved in electrospinning, predictions of the resulting fiber diameters are only possible within one protein/solvent system and under highly defined spinning conditions. Variations in the resulting fiber diameters are demonstrated in Table 16.3 giving a few examples of different electrospun protein/solvent systems.

    Allowing the production of fibers with diameters in the range of e.g. collagen fibrils (40–150 nm) (Pilotto and Filosi 1977) in combination with a high degree of biocompatibility of the underlying proteins, electrospinning of artificial silk proteins has gained rising interest for a range of applications. With regards to applications, electrospun engineered *A. diadematus* fibroin 4 (eADF4) fibers were tested as a filter material for air filtration devices, and the results showed high filter efficiencies whilst using small material amounts (Lang et al. 2013). A crucial argument for the use of silk proteins for such applications is that fragmented submicron- or nano-fibers can be respirable and harmful if they are bio-persistent. Consisting of biodegradable amino acid chains, silk proteins are assumed to be uncritical upon

Table 16.3 Examples of silk-proteins electrospun from 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), aqueous solution in the presence of 1 % w/v polyethylene oxide (PEO) or formic acid (FA). Post-treatment was performed by vaporization or immersion in methanol (MeOH)

Silk type	MW [kDa]	Solvent	c [%w/v]	d [nm]	Post-treatment	References
MaSp 1 (*N. clavipes*)	65	HFIP	15	300	–	Stephens et al. (2005)
MaSp 1 (RGD) (*N. clavipes*)	50	HFIP	2	50–250	70 % MeOH	Bini et al. (2006)
AmelF3 (*A. mellifera*)	34	aqueous 1 %w/v (PEO)	12.5	200	MeOH or water annealing	Wittmer et al. (2011)
MaSP 2 (*A. diadematus*)	48	HFIP	12	150	MeOH-vapor	Leal-Egana et al. (2012)
			16	250		
			20	480		
			24	680		
Fusion protein: silk-elastin SELP-59-A	55	FA	13	150 ± 22	MeOH-saturated air	Machado et al. (2013)
			17	183 ± 36		
			21	330 ± 40		

inhalation. Furthermore, due to their mechanical properties (predominantly soft when exposed to water), different studies have examined the suitability of electrospun recombinant silk proteins as substrates for soft-tissue cells such as fibroblasts. Engineered *A. diadematus* fibroin 4 (eADF4) proteins were electrospun using different protein concentrations in HFIP to create non-woven meshes with controllable fiber diameters in the range of 150–680 nm (Leal-Egana et al. 2012), which allowed controlling cell behavior; there was a strong correlation between the fiber diameter and the ability of mouse fibroblasts to adhere and proliferate.

Fibroblast growth was also tested on fibers made of silk-elastin-like proteins (SELPs) (70 kDa) electrospun from formic acid (15 and 20 % w/v) (Qiu et al. 2010). After posttreatment with methanol and/or glutaraldehyde vapor, water stable nonwoven meshes with fiber diameters of 50–200 nm were seeded with fibroblasts. The cells adhered to and spread on the scaffolds, maintaining high viability and proliferation rates (similar to the positive control on treated cell culture plates). Approaches to electrospin aqueous solutions were performed successfully by adding PEO and SDS to the aqueous silk-elastin like protein (70 kDa) solution, forming a polymer-micelle complex which decreases the surface tension, increases solution conductivity and retards the gelation process of SELP in water (Nagarajan et al. 2007). Furthermore, SELP-fibers could also be electrospun from pure aqueous solution at concentrations of 5–13 %w/v, though the resultant fiber mats were not as homogeneous in fiber diameter and shape as fibers electrospun from formic acid (Machado et al. 2013). Efficient cell adhesion and proliferation was also observed on electrospun fiber mats made of a recombinant honey bee larva cocoon silk protein AmelF3 (34 kDa) (Weisman et al. 2010), which was spun from aqueous solution (12.5 % w/v) (Wittmer et al. 2011). Electrospinning of 12.5 % aqueous silk solution was only possible upon the addition of 0.67 or 1.0 % w/v polyethylene oxide (PEO) and

resulted in fibers with a diameter of 200nm. Post-treatment by annealing in methanol or water caused an increase in β-sheet content and simultaneously washing out of PEO.

To further enhance their suitability for tissue engineering, recombinant silk proteins with cell binding motifs were electrospun, such as an RGD-modified *N. clavipes* MaSp 1 protein (Bini et al. 2006) or a recombinant silk-like polymer with fibronectin functionality (Anderson et al. 1994). The generated porous films showed enhanced biocompatibility based on structural gradients and controlled morphology (Buchko et al. 1999).

16.3.2   *Particles and Capsules*

Particles To produce particles from silk proteins, different procedures have been successfully applied. One strategy is to trigger self-assembly of the silk proteins, which has been demonstrated with an engineered MaSp 1 spidroin (model: *N. clavipes*) (Rabotyagova et al. 2009). Upon variation of the number of hydrophobic (alanine-rich) blocks the assembly behavior was determined by the size of the hydrophobic domain and its interaction with the solvent system. Particle formation predominantly occurred with smaller proteins in water, while micelles were formed by increasing the number of hydrophobic blocks and using isopropanol as a solvent. Furthermore, recombinant *A. trifasciata* aciniform spidroin 1 can also assemble into nanoparticles in the presence of low concentrations (near-physiological) of cosmotropic anions such as phosphate (Xu et al. 2013). This effect is likely based on the amphiphilic character of silk proteins, leading to formation of spherical structures upon reaching a critical concentration. Another approach is to 'salt-out' the proteins. This sort of protocol was applied to produce particles of an engineered *A. diadematus* fibroin 4 (eADF4) (Slotta et al. 2008). Above a critical potassium phosphate concentration, particles occurred as a result of the structural transition from random coils to β-sheets. The mechanism behind this change in structure was determined to be a liquid-liquid phase separation where there is a dense protein phase in which formation and growth of nuclei takes place. Spherical growth stagnates as the protein concentration in the protein-rich phase falls below a critical number. To control particle size, protein concentration and mixing conditions were determined as effective parameters, and particles could be produced covering a range of 250 nm- 2.1 µm (Lammel et al. 2008). In further studies, these microspheres were shown to be promising drug delivery vehicles. eADF4(C16)-particles could be efficiently loaded (almost 100 %) with lysozyme due to electrostatic interactions between the negatively charged silk proteins and positively charged lysozyme molecules (Hofer et al. 2012). Subsequent drug release could be triggered by pH and ionic strength of the release medium. Similar results were obtained with a variety of small molecular weight model drugs predominantly positively charged at pH 7 (Lammel et al. 2011). To further allow loading with negatively charged substances, a positively charged analogue of eADF4(C16), namely eADF4(κ16), was processed into particles, and

high-molecular-weight substances as well as low-molecular-weight substances could be efficiently loaded (Doblhofer and Scheibel 2015). Additionally, eADF4 proteins have been used for colloidal stabilization of hydrophobic particles in an aqueous environment (Liebmann et al. 2008). Unlike the aforementioned approaches, instead of adding the model compound (β-carotene) to the finished particles, it was added to an aqueous eADF4-solution, and subsequent salting-out was performed. It was shown that, in solution, hydrophobic regions of eADF4 interact with the hydrophobic model compound leading to a stable suspension. Salting-out resulted in coprecipitation of β-carotene encapsulated in eADF4 particles, and subsequent release could be induced by proteolytic degradation. Resistance of eADF4 particles in gastric fluids and digestion in intestinal fluids renders this drug carrier system highly interesting in the field of food and pharma formulations.

To produce particles from natural *B. mori* fibroin fibers without the use of additional chemicals as required for salting out (Lammel et al. 2010), a wet milling-spray drying process was applied (Rajkhowa et al. 2012). This top down process resulted in highly porous particles with an average diameter of 5 μm and a high surface area. Furthermore, it was shown, that milling induced breaking of stacked β-sheets within silk crystallites leading to a significant increase of enzymatic hydrolysis compared to that of fibers. Such tuning of the degradation behavior of porous silk particles can be attractive in developing targeted biomedical applications.

Capsules Formation of a continuous protein membrane is a prerequisite to produce capsules, and there are different strategies to obtain capsules of proteins. Self-assembly of amphiphilic proteins and surfactant peptides has been shown to be sufficient for smaller molecules (Park and Champion 2014; Vauthey et al. 2002; Santoso et al. 2002), but is likely unsuitable for more complex larger proteins with more functionalities such as silk proteins (Hermanson et al. 2007b). Other approaches apply micro-spherical templates providing interfaces at which protein layers can form either by electrostatic deposition (Du et al. 2014; Zhao and Li 2008; Li et al. 2005) or by adsorption of amphiphilic protein chains due to hydrophobic interactions (Horinek et al. 2008; Geisler et al. 2008).

Regenerated *B. mori* silk fibroin has been processed into microcapsules by electrostatically driven layer-by-layer deposition on polystyrene particles as a sacrificial template (Li et al. 2014). The particles were first coated with polyethylenimine to render their surface positively charged and subsequently coated with up to 10 silk fibroin layers. The polystyrene cores were dissolved in tetrahydrofuran to obtain microcapsules that could be shown to be potentially suitable as carriers for controlled, localized gene delivery. Alternatively, silica particles can be coated with aqueous silk solution and subsequently removed by dissolution in 8 % hydrogen fluoride (Shchepelina et al. 2011). The resulting capsules could potentially be applied as bio-sensing materials.

In an emulsion-based approach, a recombinantly produce eADF4 spider silk protein was shown to assemble at an oil/water interface to form mechanically stable

microcapsules (Hermanson et al. 2007b). Emulsification of an aqueous eADF4 solution in toluene resulted in formation of microcapsules due to diffusion-driven adsorption, assembly and β-sheet formation of silk proteins at the water/toluene interface. The size of the capsules (1–30 μm) could be controlled by shear-rate adjusted droplet-size of the emulsion. The molecular cut-off of the porous capsules was measured to be 27 kDa enabling small molecules to freely diffuse through the capsule. To induce further release of bigger molecules, proteinase K can be incorporated to destroy the membrane's integrity. The determined cut-off was assumed to provide good protection against the immunological system, rendering it interesting for applications such as enzyme therapy and artificial cells (Hermanson et al. 2007a). In a further study, toluene was replaced by medical grade silicon oil to avoid toxic effects, and protein encapsulation was performed using β-galactosidase as a model enzyme (Blum et al. 2014). Using silicon oil, an additional incubation step in ethanol was required to induce a stable β-sheet-rich secondary structure. Catalytic activity of enzymes was visualized by conversion of a colorless substrate into its colored product: the substrate was provided outside of the capsules and freely diffused in. Furthermore, an encapsulated inactive variant of β-galactosidase could be activated by diffusion of an activation peptide into the silk capsules. In the presence of distinct proteases, such as AspN, it could be shown that the silk capsule protected β-galactosidase from proteolytic degradation. The combination of good mechanical stability (Hermanson et al. 2007b), biocompatibility and controllable degradation render spider silk capsules highly interesting as enclosed enzyme reaction chambers for various drug delivery or diagnostic sensor systems.

16.3.3 Hydrogels, Foams and Sponges

Hydrogels Hydrogels display attractive properties, especially in the field of tissue regeneration, as they exhibit characteristics similar to that of soft tissues (water-swollen network, soft mechanics). Additionally, they are attractive for 3D cell-culture due to efficient mass transfer as well as biocompatible gelation processes (Zhu and Marchant 2011). Various synthetic and naturally derived materials have been processed into hydrogels and tested for their medical suitability (Drury and Mooney 2003). Owing to their mechanical properties, biocompatibility and constrained degradation behavior, also silk materials are considered a promising class of materials in this field (Jonker et al. 2012).

The feasibility of creating hydrogels from silk and their suitability for tissue engineering was demonstrated using regenerated *B. mori* silk fibroin. To produce silk fibroin hydrogels, gelation can be induced either by sonification (Etienne et al. 2009) or by vortexing (Yucel et al. 2009) of aqueous protein solutions. The gelation mechanism was described as a result of shear-induced amphiphilic cluster association due to an increase of concentration fluctuations in aqueous solution, serving as a poor solvent system (Yucel et al. 2009). *B. mori* silk hydrogels were unique in their capability to provide a neutral and hydrophilic environment combined with

sufficient mechanical properties to support e.g. the spherical chondrocyte phenotype found in healthy cartilage (Chao et al. 2010).

Aqueous solutions of a recombinantly produced eADF4 spider silk protein were shown to spontaneously form hydrogels due to a nucleation-aggregation mechanism resulting in a nanofibrillar hydrogel network (Slotta et al. 2007). The gelation process is accompanied by the transition of amorphous conformations into β-sheet-rich ones, where protein and ion concentration as well as ion composition play an important role. Dialysis of low concentrated protein solutions was applied to obtain a regulated increase in protein concentration allowing the removal of the solution prior to hydrogel formation (Schacht and Scheibel 2011). With regards to potential tissue engineering applications, the latter is crucial to enable addition of cells into the solution and subsequent encapsulation by gelation (Schacht et al. 2015). To enhance cell adhesion in the resulting hydrogel, an RGD-modified variant of eADF4 was used. Further studies on hydrogel formation showed that variations in protein concentration (30–70 mg/ml) can be used to control the pore size (200–30 μm) and elastic modulus (0.1–35 kPa). Chemical crosslinking of eADF4-hydrogels with ammonium peroxodisulphate and tris(2,2'-bipyridyl)dichlororuthenium(II) resulted in stabilized morphologies with increased elastic modulus (110 kPa) (Schacht and Scheibel 2011). Adjustable pore sizes and mechanical properties are considered important prerequisites to optimize hydrogels for distinct applications. Towards applications in the medical field, eADF4 hydrogels display a shear thinning behavior, which allows their use as a bioink for conventional 3D-bioplotting to create well defined 3D scaffolds (Schacht et al. 2015). Strikingly, eADF4 shows pronounced printing fidelity and can be plotted without the need of additives or post-processing procedures, unlike many other conventional bioinks including silkworm fibroin (Malda et al. 2013, Das et al. 2015). Moreover, cells could be printed when embedded in the hydrogel, and cell viability was evidenced for 7 days, a behaviour attributed to the material's biocompatibility and diffusivity.

Due to their highly tunable pore size and gelation properties, recombinant protein hydrogels can also be applied as drug delivery systems as shown for silk-elastin like proteins (Dinerman et al. 2002; Megeed et al. 2002). A hybrid with combined silk and elastin amino acid sequences was shown to be temperature-sensitive and irreversibly formed hydrogels by β-sheet formation of the silk motive when exposed to 37 °C (Gustafson et al. 2010). This behavior enables *in vivo* gelation of a drug-loaded silk solution. To test SELP hydrogels for their suitability in cancer therapy, *in vivo* experiments (model: mouse) were performed by intra-tumoral injection of hydrogels which were loaded with adenoviruses frequently used in gene therapy. Different studies demonstrated the ability of SELP hydrogels to mediate the delivery of adenoviruses reducing the systemic toxicity of free viral injection (in saline) applied in gene-directed enzyme-prodrug therapy (Gustafson et al. 2010; Greish et al. 2010; Megeed et al. 2004). The underlying diffusion and release kinetics of such hydrogels were explicitly studied regarding solute hydrophobicity and charge (Dinerman et al. 2010). The results indicated extended control of the solute release from SELP hydrogels by modification of the hydrophobicity and charge of drugs.

Foams and Sponges Foams and sponges are highly porous, 3-dimensional scaffolds which are frequently applied for tissue engineering. In this context, they serve as a physical surface which can be a substrate for the generation of an endogenous extracellular matrix (ECM) by the cells, as well as allow for efficient mass transport due to the relatively large pore sizes. Furthermore, cell growth of adherent contact-inhibited cells may be prevented, and potential cluster sizes are reduced to that of the pore size, excluding the growth of large cell clusters that might form a necrotic center (Dhandayuthapani et al. 2011). Different methods to produce foams and sponges have been reported such as phase-separation or the emulsion freeze-drying method, both of which are based on freeze-drying of a biphasic system. Another technique, the porogen leaching method, is often favored in tissue engineering as it provides superior control over pore size and structure (Chen et al. 2002). Thereby, a homogeneous suspension of defined salt particles is created in a protein or polymer solution, which is then cast and dried. The composite is subsequently immersed in water to leach out the salt particles leaving a porous structure (Mikos et al. 1994).

The latter was evaluated in the field of cartilage and bone repair applying *B. mori* silk fibroin as a scaffold material prepared in HFIP with NaCl as a porogen, where resulting scaffolds had a porosity of ~95.5 % (Marolt et al. 2006). Post-treatment was performed by immersion in methanol to induce a water-stable β-sheet rich silk structure. Subsequent cell culture experiments indicated appropriate properties of such scaffolds for bone regeneration. In an alternative approach, silk fibroin was dissolved in 9.3 M lithium bromide solution, which was advantageous as no post-treatment was required (Wang et al. 2005; Bhardwaj et al. 2011). The resultant scaffolds displayed a pore size of 550 ± 30 μm, and subsequent cell culture experiments showed promising results concerning cell-based cartilage tissue engineering.

Recently, NaCl salt-leaching was applied to produce scaffolds made of the recombinant eADF4 spider silk protein (Schacht et al. 2016). The study included a broad range of processing conditions, including various protein concentrations (4, 6, 8 % w/v), salt-to-silk ratio, and NaCl crystal sizes (45–158 μm, 121–320 μm, 319–514 μm). It could be shown that HFIP-derived foams exhibit high β-sheet contents owing to the presence of NaCl crystals acting as structure-conversion nuclei for β-sheet formation. Thus, unlike the similarly salt-leached *B. mori fibroin* scaffolds described above (Marolt et al. 2006), no additional post-treatment was necessary for eADF4 foams. Furthermore, it was shown that the mechanical properties of the scaffolds could be adjusted by the protein concentration covering compressive moduli within the range of soft tissues from 0.94 ± 0.26 kPa (4 % w/v) to 3.24 ± 1.03 kPa (8 % w/v). This is important, since matrix elasticity is assumed to have a high impact on differentiation behavior of human mesenchymal stem cells. Cell culture experiments with RGD-modified eADF4 foams showed the best result for scaffolds with middle pore sizes (salt crystals: 121–320 μm, resulting pores: 59–248 μm). This effect could be attributed to the well-balanced combination of sufficient nutrient/waste diffusion and the large available surface area. Similar observations were obtained for scaffolds made of recombinantly produced spidroin 1 (model: *N. clavipes*) (Agapov et al. 2009). Porogen leaching was performed using 10 % lithium chloride and 90 % formic acid as a solvent system employing NaCl crystals (50–

100 μm and 200–400 μm) as sacrificial particles. Cell culture experiments with fibroblasts showed cell migration from the outer to the inner regions of the scaffold over a period of 14 days. The results confirm the legitimacy of the aspired use of artificial silk scaffolds in medical applications.

Another method to produce silk sponges is based on freezing and thawing of aqueous protein solutions. For *B. mori* fibroin processing, dimethyl sulfoxide (DMSO) was added (0.05–2 % v/v) to aqueous fibroin solutions to induce transformation into the β-sheet-rich conformation and thus to initiate nucleic self-assembly (Tamada 2005). The solution was subsequently frozen at −20 °C with a freezing duration of >6 h to provide enough time for network formation. After thawing of the matrix-embedded ice crystals a porous structure remained. The resulting pore size and the mechanical properties were shown to be highly dependent on the fibroin concentration in solution. A similar approach was applied to produce 3D scaffolds of a recombinant spider silk protein (rSSp) upon freezing a hydrogel at −20 °C and subsequently thawing it (Jones et al. 2015). The resulting sponges were demonstrated to exhibit shape memory enabling them to repeatedly recover after compression.

16.3.4   Films and Coatings

Films Films can be produced from different types of silk solutions by casting and solvent evaporation. Although the method itself is simple, several significant input parameters lead to various film properties.

To produce mechanically stable films of recombinant MaSp1 and MaSp2 several different parameters have been examined: Solvent systems and additives, protein concentration, casting and drying parameters, post-stretching and chemical post-treatment (Tucker et al. 2014). Morphologically, it could be seen that films cast from HFIP displayed pores, which was attributed to the evaporation rate of the highly volatile solvent. In contrast, films cast from aqueous protein solutions had no pores and mechanically matched or even outperformed those cast from HFIP. Increase in stress and strain could be accomplished by addition of glutaraldehyde which serves as a cross-linking agent. Furthermore, enhancement of film extensibility was achieved by introduction of hygroscopic plasticizers like glycerol, altering intermolecular interactions (Spiess et al. 2010). Analogous to fiber post-stretching, post-stretching while immersed in an alcohol/water mixture was applied and improved the overall mechanical properties. The impact of the initial solvent on the resulting film properties was also investigated for eADF4 proteins (Spiess et al. 2011). Using HFIP, formic acid (FA) and aqueous solutions, significant differences were reported for the resulting secondary structures and the related mechanical properties. Cast films out of HFIP displayed lower crystalline fractions (21 %) compared to the ones out of FA (32 %), whereas films cast out of aqueous solution showed structural inhomogeneity (25 % and 30 %). This effect might be attributed to inhomogeneous evaporation of water indicating the importance of well-defined

film curing conditions. Though post-treatment with methanol resulted in similar crystallinities for all film types, the mechanical properties did not assimilate. This can be explained as an artifact of solvent-dependent arrangement of β-strands emphasizing the impact of the solvent on the film properties. Furthermore, it has to be considered, that post-treatment might mainly affect the outer film layer, which is measured by FT-IR, and be less efficient in the inner part of the film. However, such conformational differences can be utilized to create functional materials by selectively patterning the secondary structure of films (Young et al. 2012).

Two lithographic methods, the capillary transfer lithography technique (CTL) and solvent-assisted micro-contact molding (SAMIM) have been applied to selectively mask eADF4-films with predominantly amorphous secondary structures. Subsequent treatment with methanol-vapor and removal of the mask resulted in precise patterning of protein secondary structure with sub-micrometer periodicity. Patterned films were also created using eADF4 proteins in combination with a recombinant lacewing egg stalk protein (Bauer et al. 2013). The films were composed of a silk ground layer with ridges (width: 50 μm, height: <1 μm, spacing: 20 μm) made of another protein, respectively. Cell culture experiments with fibroblasts and myoblast demonstrated the beneficial impact of the films' morphological features enhancing cell adhesion and alignment.

Besides their frequently demonstrated biocompatibility and mechanical strength, artificial silk protein films were also shown to have attractive optical properties (Teng et al. 2011). Films were cast from aqueous solutions of an engineered silk-elastinlike protein, and β-sheet formation was induced by immersion in methanol. It could be shown that post-treated films were nearly optically transparent in their hydrated state rendering them a potential material for ophthalmic applications such as contact lenses, synthetic corneas or intraocular lenses.

Coatings Due to their unique properties, silk proteins are considered a highly attractive coating material in the field of biomedical applications such as implant coatings (Zeplin et al. 2014a), bioimaging devices (Khalid et al. 2014) and immunosensors (Moraes et al. 2013). To generate protein coatings, dip coating (Gill and Prausnitz 2007; Lee et al. 2009), spin coating (Genov et al. 2011; Raphel et al. 2012; Junghans et al. 2006), spray coating (Maa and Hsu 1997) and electrodeposition (Elia et al. 2015; Bhardwaj and Webster 2015) have been previously applied.

Spin coating was shown to cause shear-induced conformational changes of native silk fibroins (Greving et al. 2012). Depending on the concentration of the protein solution, self-assembly into oriented nanofibrils (higher concentrations) and randomly coiled thin strings (lower concentration) was observed. Furthermore, the hydrophobicity of the coated surface can have a strong impact on the molecular assembly of a thin protein layer as described for the recombinant spider silk protein eADF4 (Wohlrab et al. 2012). Structural analysis and contact angle measurements led to the assumption that hydrophilic templates induce formation of hydrophobic silk patches embedded in the hydrophilic blocks of the amphiphilic proteins. At the

surface of the coating, hydrophobic patches are preferentially exposed to air in contrast to the hydrophilic bulk rendering the coating's surface apparently hydrophobic. Consequently, hydrophobic templates resulted in hydrophilic silk coating surfaces.

In an application-oriented approach, highly hydrophobic silicone implants were dip-coated with aqueous solutions of the recombinant spider silk protein eADF4 (Zeplin et al. 2014b). Silicone implants are resistant against hydrolytic and enzymatic degradation, have a hydrophobic surface and facilitate unspecific protein adhesion, attracting inflammatory and pro-fibrotic cells which cause body-associated fibrosis. *In vivo*-experiments (model: rat) showed, that undesired fibrotic capsule formation could be significantly reduced by masking the implant with a thin layer (~1–6 μm) of eADF4 proteins. This was attributed to the reduced cell adhesion and proliferation on flat eADF4-surfaces (Leal-Egana et al. 2012) accompanied with poor protein adsorption. The results indicate that eADF4 can be employed to avoid inflammatory and fibrotic complications, particularly from the innate immune response. Furthermore, cytocompatibility studies with other silk coatings such as plant-derived recombinant *N. clavipes* proteins support the assumption that spidroin-derived materials do not cause observable immunogenic or cytotoxicity effects (Hauptmann et al. 2015). Apart from the commonly observed good biocompatibility of silk coatings, cell adhesion- and proliferation behavior is assumed to be highly dependent on the cell type and the applied proteins. This can be seen when e.g. comparing the low fibroblast growth on eADF4 protein films (Leal-Egana et al. 2012) with the good chondrocyte growth on recombinant spider silk-elastin protein coated tissue culture plates (Scheller et al. 2004). The latter study showed that spider silk-elastin films enhanced chondrocyte proliferation rates and are therefore regarded a potential alternative for collagen-coated cell culture plates.

16.4   Conclusion and Outlook

Silk proteins can be processed into a range of morphologies which reflect a highly attractive class of multipurpose materials displaying a unique variety of application-relevant properties. Though silk has to compete with significantly cheaper synthetic polymers, the combination of mechanical properties, biocompatibility, environmental sustainability, ease of functionalization, good optical properties and morphological diversity trigger increasing industrial interest. Useful applications generally require high performance materials with specific properties as demanded, for example in medical devices. The most intensively studied silk materials are based on regenerated natural *B. mori* fibroin due to its availability, but also on recombinant spider silk proteins due to their good physico-chemical and biomedical properties. Though two decades have passed since the first successful attempts were made to recombinantly produce spider silk (Prince et al. 1995; Arcidiacono et al. 1998), industrial implementation is still in its infancy. Taking into account that the most promising applications might be in biomedicine, it is not surprising that getting new materials based on recombinant silk proteins ready for such applications is very time consuming.

References

Addison JB, Ashton NN, Weber WS, Stewart RJ, Holland GP, Yarger JL (2013) beta-Sheet nano-crystalline domains formed from phosphorylated serine-rich motifs in caddisfly larval silk: a solid state NMR and XRD study. Biomacromolecules 14:1140–1148

Addison JB, Weber WS, Mou Q, Ashton NN, Stewart RJ, Holland GP, Yarger JL (2014) Reversible assembly of beta-sheet nanocrystals within caddisfly silk. Biomacromolecules 15:1269–1275

Adrianos SL, Teule F, Hinman MB, Jones JA, Weber WS, Yarger JL, Lewis RV (2013) *Nephila clavipes* Flagelliform silk-like GGX motifs contribute to extensibility and spacer motifs contribute to strength in synthetic spider silk fibers. Biomacromolecules 14:1751–1760

Agapov II, Pustovalova OL, Moisenovich MM, Bogush VG, Sokolova OS, Sevastyanov VI, Debabov VG, Kirpichnikov MP (2009) Three-dimensional scaffold made from recombinant spider Silk protein for tissue engineering. Dokl Biochem Biophys 426:127–130

Albertson AE, Teule F, Weber W, Yarger JL, Lewis RV (2014) Effects of different post-spin stretching conditions on the mechanical properties of synthetic spider silk fibers. J Mech Behav Biomed Mater 29:225–234

Altman GH, Diaz F, JAKUBA C, Calabro T, Horan RL, Chen J, Lu H, Richmond J, Kaplan DL (2003) Silk-based biomaterials. Biomaterials 24:401–416

An B, Hinman MB, Holland GP, Yarger JL, Lewis RV (2011) Inducing beta-sheets formation in synthetic spider silk fibers by aqueous post-spin stretching. Biomacromolecules 12:2375–2381

Anderson JP, Cappello J, Martin DC (1994) Morphology and primary crystal structure of a silk-like protein polymer synthesized by genetically engineered *Escherichia coli* bacteria. Biopolymers 34:1049–1058

Arcidiacono S, Mello C, Kaplan D, Cheley S, Bayley H (1998) Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. Appl Microbiol Biotechnol 49:31–38

Ashton NN, Stewart RJ (2015) Self-recovering caddisfly silk: energy dissipating, Ca(2+)-dependent, double dynamic network fibers. Soft Matter 11:1667–1676

Ashton NN, Roe DR, Weiss RB, Cheatham TE, Stewart RJ (2013) Self-tensioning aquatic caddis-fly silk: Ca2+−dependent structure, strength, and load cycle hysteresis. Biomacromolecules 14:3668–3681

Askarieh G, Hedhammar M, Nordling K, Saenz A, Casals C, Rising A, Johansson J, Knight SD (2010) Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. Nature 465:236–238

Atkins EDT (1967) A four-strand coiled coil model for some insect fibrous proteins. J Mol Biol 24:139–141

Ayoub NA, Garb JE, Tinghitella RM, Collin MA, Hayashi CY (2007) Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. PLoS One 2:e514

Bai X, Sakaguchi M, Yamaguchi Y, Ishihara S, Tsukada M, Hirabayashi K, Ohkawa K, Nomura T, Arai R (2015) Molecular cloning, gene expression analysis, and recombinant protein expression of novel silk proteins from larvae of a retreat-maker caddisfly, *Stenopsyche marmorata*. Biochem Biophys Res Commun 464:814–819

Barr LA, Fahnestock SR, Yang J (2004) Production and purification of recombinant DP1B silk-like protein in plants. Mol Breed 13:345–356

Bauer F, Bertinetti L, Masic A, Scheibel T (2012) Dependence of mechanical properties of lace-wing egg stalks on relative humidity. Biomacromolecules 13:3730–3735

Bauer F, Scheibel T (2012) Artificial egg stalks made of a recombinantly produced lacewing silk protein. Angew Chem 51:6521–6524

Bauer F, Wohlrab S, Scheibel T (2013) Controllable cell adhesion, growth and orientation on layered silk protein films. Biomed Sci 1:1244–1249

Becker N, Oroudjev E, Mutz S, Cleveland JP, Hansma PK, Hayashi CY, Makarov DE, Hansma HG (2003) Molecular nanosprings in spider capture-silk threads. Nat Mater 2:278–283

Bhardwaj G, Webster T (2015) Coating polyurethane surfaces by electrostatic charging followed by dip coating/electrophoretic deposition. FASEB J:29

Bhardwaj N, Nguyen QT, Chen AC, Kaplan DL, Sah RL, Kundu SC (2011) Potential of 3-D tissue constructs engineered from bovine chondrocytes/silk fibroin-chitosan for in vitro cartilage tissue engineering. Biomaterials 32:5773–5781

Bini E, Knight DP, Kaplan DL (2004) Mapping domain structures in silks from insects and spiders related to protein assembly. J Mol Biol 335:27–40

Bini E, Foo CWP, Huang J, Karageorgiou V, Kitchel B, Kaplan DL (2006) RGD-functionalized bioengineered spider dragline silk biomaterial. Biomacromolecules 7:3139–3145

Blackledge TA, Summers AP, Hayashi CY (2005) Gumfooted lines in black widow cobwebs and the mechanical properties of spider capture silk. Zoology 108:41–46

Blüm C, Scheibel T (2012) Control of drug loading and release properties of spider silk submicroparticles. Bio Nano Science 2:67–74

Blum C, Nichtl A, Scheibel T (2014) Spider silk capsules as protective reaction containers for enzymes. Adv Funct Mater 24:763–768

Borkner CB, Elsner MB, Scheibel T (2014) Coatings and films made of silk proteins. ACS Appl Mater Interfaces 6:15611–15625

Brubaker CE, Messersmith PB (2012) The present and future of biologically inspired adhesive interfaces and materials. Langmuir : the ACS journal of surfaces and colloids 28:2200–2205

Buchko CJ, Chen LC, Shen Y, Martin DC (1999) Processing and microstructural characterization of porous biocompatible protein polymer thin films. Polymer 40:7397–7407

Cebe P, Hu X, Kaplan DL, Zhuravlev E, Wurm A, Arbeiter D, Schick C (2013) Beating the heat - fast scanning melts silk beta sheet crystals. Sci Rep 3

Cereghino GP, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris. Curr Opin Biotechnol 13:329–332

Challis RJ, Goodacre SL, Hewitt GM (2006) Evolution of spider silks: conservation and diversification of the C-terminus. Insect Mol Biol 15:45–56

Chao PHG, Yodmuang S, Wang XQ, Sun L, Kaplan DL, Vunjak-Novakovic G (2010) Silk hydrogel for cartilage tissue engineering. J Biomed Mater Res B Appl Biomater 95B:84–90

Chen GP, Ushida T, Tateishi T (2002) Scaffold design for tissue engineering. Macromol Biosci 2:67–77

Chen X, Cai H, Ling S, Shao Z, Huang Y (2012) Conformation transition of *Bombyx mori* silk protein monitored by time-dependent fourier transform infrared (FT-IR) spectroscopy: effect of organic solvent. Appl Spectrosc 66:696–699

Craig CL (1997) Evolution of arthropod silks. Annu Rev Entomol 42:231–267

Cregg JM, Vedvick TS, Raschke WC (1993) Recent advances in the expression of foreign genes in Pichia pastoris. Bio/Technology 11:905–910

Das S, Pati F, Choi YJ, Rijal G, Shim JH, Kim SW, Ray AR, Cho DW, Ghosh S (2015) Bioprintable, cell-laden silk fibroin-gelatin hydrogel supporting multilineage differentiation of stem cells for fabrication of three-dimensional tissue constructs. Acta Biomater 11:233–246

Dhandayuthapani B, Yoshida Y, Maekawa T, Kumar DS (2011). Polymeric scaffolds in tissue engineering application: a review. Int J Polym Sci 2011:1–9

Dicko C, Knight D, Kenney JM, Vollrath F (2004) Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects. Biomacromolecules 5:2105–2115

Dinerman AA, Cappello J, El-Sayed M, Hoag SW, Ghandehari H (2010) Influence of solute charge and hydrophobicity on partitioning and diffusion in a genetically engineered silk-elastin-like protein polymer hydrogel. Macromol Biosci 10:1235–1247

Dinerman AA, Cappello J, Ghandehari H, Hoag SW (2002) Swelling behavior of a genetically engineered silk-elastinlike protein polymer hydrogel. Biomaterials 23:4203–4210

Doblhofer E, Scheibel T (2015) Engineering of recombinant spider silk proteins allows defined uptake and release of substances. J Pharm Sci 104:988–994

Drury JL, Mooney DJ (2003) Hydrogels for tissue engineering: scaffold design variables and applications. Biomaterials 24:4337–4351

Du B, Wang JJ, Zhou ZM, Tang HB, LI XM, Liu YJ, Zhang QQ (2014) Synthesis of silk-based microcapsules by desolvation and hybridization. Chem Commun 50:4423–4426

Elia R, Michelson CD, Perera AL, Brunner TF, Harsono M, Leisk GG, Kugel G, Kaplan DL (2015) Electrodeposited silk coatings for bone implants. J Biomed Mater Res B Appl Biomater 103:1602–1609

Engster MS (1976) Studies on silk secretion in the trichoptera (F. Limnephilidae). Cell Tissue Res 169:77–92

Etienne O, Schneider A, Kluge JA, Bellemin-Laponnaz C, Polidori C, Leisk GG, Kaplan DL, Garlick JA, Egles C (2009) Soft tissue augmentation using silk gels: an in vitro and in vivo study. J Periodontol 80:1852–1858

Fahnestock SR, Bedzyk LA (1997) Production of synthetic spider dragline silk protein in Pichia pastoris. Appl Microbiol Biotechnol 47:33–39

Garb JE, Ayoub NA, Hayashi CY (2010) Untangling spider silk evolution with spidroin terminal domains. BMC Evol Biol 10:243

Garb JE, Hayashi CY (2005) Modular evolution of egg case silk genes across orb-weaving spider superfamilies. *Proc Natl Acad Sci U S A* 102:11379–11384

Garg K, Bowlin GL (2011) Electrospinning jets and nanofibrous structures. Biomicrofluidics 5

Geisler M, Pirzer T, Ackerschott C, Lud S, Garrido J, Scheibel T, Hugel T (2008) Hydrophobic and Hofmeister effects on the adhesion of spider silk proteins onto solid substrates: an AFM-based single-molecule study. Langmuir 24:1350–1355

Genov S, Riester D, Hirth T, Tovar G, Borchers K, Weber A (2011) Preparation and characterisation of dry thin native protein trehalose films on titanium-coated cyclo-olefin polymer (COP) foil generated by spin-coating/drying process and applied for protein transfer by Laser-Induced-Forward Transfer (LIFT). Chem Eng Process 50:558–564

Gill HS, Prausnitz MR (2007) Coating formulations for microneedles. Pharm Res 24:1369–1380

Gosline JM, Denny MW, Demont ME (1984) Spider silk as rubber. Nature 309:551–552

Gosline JM, Guerette PA, Ortlepp CS, Savage KN (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function. J Exp Biol 202:3295–3303

Gotoh Y, Tsukada M, Baba T, Minoura N (1997) Physical properties and structure of poly(ethylene glycol)-silk fibroin conjugate films. Polymer 38:487–490

Greiner A, Wendorff JH (2007) Electrospinning: a fascinating method for the preparation of ultra-thin fibres. Angew Chem Int Ed 46:5670–5703

Greish K, Frandsen J, Scharff S, Gustafson J, Cappello J, Li DQ, O'malley BW, Ghandehari H (2010) Silk-elastinlike protein polymers improve the efficacy of adenovirus thymidine kinase enzyme prodrug therapy of head and neck tumors. J Gene Med 12:572–579

Greving I, Cai MZ, Vollrath F, Schniepp HC (2012) Shear-induced self-assembly of native silk proteins into fibrils studied by atomic force microscopy. Biomacromolecules 13:676–682

Grip S, Rising A, Nimmervoll H, Storckenfeldt E, Mcqueen-Mason SJ, Pouchkina-Stantcheva N, Vollrath F, Engsträm W, Fernandez-Arias A (2006) Transient expression of a major ampullate spidroin 1 gene fragment from *Euprosthenops sp*. in mammalian cells. Cancer Genomics Proteomics 3:83–87

Guerette PA, Ginzinger DG, Weber BH, Gosline JM (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family. Science 272:112–115

Gustafson JA, Price RA, Greish K, Cappello J, Ghandehari H (2010) Silk-elastin-like hydrogel improves the safety of adenovirus-mediated gene-directed enzyme-prodrug therapy. Mol Pharm 7:1050–1056

Hagn F, Eisoldt L, Hardy JG, Vendrely C, Coles M, Scheibel T, Kessler H (2010) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. Nature 465:239–242

Hagn F, Thamm C, Scheibel T, Kessler H (2011) pH-dependent dimerization and salt-dependent stabilization of the N-terminal domain of spider dragline silk–implications for fiber formation. Angew Chem 50:310–313

Hajer J, Rehakova D (2003) Spinning activity of the spider *Trogloneta granulum* (Araneae, Mysmenidae): web, cocoon, cocoon handling behaviour, draglines and attachment discs. Zoology 106:223–231

Hauptmann V, Weichert N, Menzel M, Knoch D, Paege N, Scheller J, Spohn U, Conrad U, Gils M (2013) Native-sized spider silk proteins synthesized in planta via intein-based multimerization. Transgenic Res 22:369–377

Hauptmann V, Menzel M, Weichert N, Reimers K, Spohn U, Conrad U (2015) In planta production of ELPylated spidroin-based proteins results in non-cytotoxic biopolymers. BMC Biotechnol 15:9

Hayashi CY, Blackledge TA, Lewis RV (2004) Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Mol Biol Evol 21:1950–1959

Heidebrecht A, Scheibel T (2013) Recombinant production of spider silk proteins. Adv Appl Microbiol 82:115–153

Heidebrecht A, Eisoldt L, Diehl J, Schmidt A, Geffers M, Lang G, Scheibel T (2015) Biomimetic fibers made of recombinant spidroins with the same toughness as natural spider silk. Adv Mater 27:2189–2194

Heim M, Keerl D, Scheibel T (2009) Spider silk: from soluble protein to extraordinary fiber. Angew Chem 48:3584–3596

Heim M, Romer L, Scheibel T (2010) Hierarchical structures made of proteins. The complex architecture of spider webs and their constituent silk proteins. *Chem Soc Rev* 39:156–164

Hepburn HR, Chandler HD, Davidoff MR (1979) Extensometric properties of insect fibroins – green lacewing cross-beta, honeybee alpha-helical and greater waxmoth parallel-beta conformations. Insect Biochem 9:69–77

Hepburn HR, Kurstjens SP (1988) The combs of honeybees as composite materials. Apidologie 19:25–36

Hermanson KD, Harasim MB, Scheibel T, Bausch AR (2007a) Permeability of silk microcapsules made by the interfacial adsorption of protein. Phys Chem Chem Phys 9:6442–6446

Hermanson KD, Huemmerich D, Scheibel T, Bausch AR (2007b) Engineered microcapsules fabricated from reconstituted spider silk. *Adv Mater* 19:1810

Hijirida DH, Do KG, Michal C, Wong S, Zax D, Jelinski LW (1996) 13C NMR of *Nephila clavipes* major ampullate silk gland. Biophys J 71:3442–3447

Hofer M, Winter G, Myschik J (2012) Recombinant spider silk particles for controlled delivery of protein drugs. Biomaterials 33:1554–1562

Holland GP, Jenkins JE, Creager MS, Lewis RV, Yarger JL (2008) Quantifying the fraction of glycine and alanine in beta-sheet and helical conformations in spider dragline silk using solid-state NMR. Chem Commun 43:5568–5570

Horinek D, Serr A, Geisler M, Pirzer T, Slotta U, Lud SQ, Garrido JA, Scheibel T, Hugel T, Netz RR (2008) Peptide adsorption on a hydrophobic surface results from an interplay of solvation, surface, and intrapeptide forces. *Proc Natl Acad Sci U S A* 105:2842–2847

Hu X, Yuan J, Wang X, Vasanthavada K, Falick AM, Jones PR, La Mattina C, Vierra CA (2007) Analysis of aqueous glue coating proteins on the silk fibers of the cob weaver, *Latrodectus hesperus*. Biochemistry 46:3294–3303

Huemmerich D, Helsen CW, Quedzuweit S, Oschmann J, Rudolph R, Scheibel T (2004a) Primary structure elements of spider dragline silks and their contribution to protein solubility. Biochemistry 43:13604–13612

Huemmerich D, Scheibel T, Vollrath F, Cohen S, Gat U, Ittah S (2004b) Novel assembly properties of recombinant spider dragline silk proteins. Curr Biol 14:2070–2074

Inoue S, Tanaka K, Arisaka F, Kimura S, Ohtomo K, Mizuno S (2000) Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio. J Biol Chem 275:40517–40528

Jenkins JE, Creager MS, Butler EB, Lewis RV, Yarger JL, Holland GP (2010) Solid-state NMR evidence for elastin-like beta-turn structure in spider dragline silk. Chem Commun 46:6714–6716

Johansson J, Nerelius C, Willander H, Presto J (2010) Conformational preferences of non-polar amino acid residues: an additional factor in amyloid formation. Biochem Biophys Res Commun 402:515–518

Jones JA, Harris TI, Tucker CL, Berg KR, Christy SY, Day BA, Gaztambide DA, Needham NJ, Ruben AL, Oliveira PF, Decker RE, Lewis RV (2015) More than just fibers: an aqueous method for the production of innovative recombinant spider silk protein materials. Biomacromolecules 16:1418–1425

Jonker AM, Lowik DWPM, Van Hest JCM (2012) Peptide- and protein-based hydrogels. Chem Mater 24:759–773

Junghans F, Morawietz M, Conrad U, Scheibel T, Heilmann A, Spohn U (2006) Preparation and mechanical properties of layers made of recombinant spider silk proteins and silk from silk worm. Appl Phys Mater Sci Process 82:253–260

Kameda T, Walker AA, Sutherland TD (2014) Evolution and application of coiled coil silks from insects. In: Asakura T, Miller T (eds) Biotechnology of silk. Springer, Dordrecht

Kamenskiy AV, Dzenis YA, Kazmi SAJ, Pemberton MA, Pipinos II, Phillips NY, Herber K, Woodford T, Bowen RE, Lomneth CS, Mactaggart JN (2014) Biaxial mechanical properties of the human thoracic and abdominal aorta, common carotid, subclavian, renal and common iliac arteries. Biomech Model Mechanobiol 13:1341–1359

Khalid A, Lodin R, Domachuk P, Tao H, Moreau JE, Kaplan DL, Omenetto FG, Gibson BC, Tomljenovic-HANIC S (2014) Synthesis and characterization of biocompatible nanodiamond-silk hybrid material. Biomed Opt Express 5:596–608

Kinahan ME, Filippidi E, Koster S, Hu X, Evans HM, Pfohl T, Kaplan DL, Wong J (2011) Tunable silk: using microfluidics to fabricate silk fibers with controllable properties. Biomacromolecules 12:1504–1511

Lammel A, Schwab M, Slotta U, Winter G, Scheibel T (2008) Processing conditions for the formation of spider silk microspheres. ChemSusChem 1:413–416

Lammel AS, Hu X, Park SH, Kaplan DL, Scheibel TR (2010) Controlling silk fibroin particle features for drug delivery. Biomaterials 31:4583–4591

Lammel A, Schwab M, Hofer M, Winter G, Scheibel T (2011) Recombinant spider silk particles as drug delivery vehicles. Biomaterials 32:2233–2240

Lane DD, Kaur S, Weerasakare GM, Stewart RJ (2015) Toughened hydrogels inspired by aquatic caddisworm silk. Soft Matter 11:6981–6990

Lang G, Jokisch S, Scheibel T (2013) Air filter devices including nonwoven meshes of electrospun recombinant spider silk proteins. J Vis Exp 75:e50492

Lazaris A, Arcidiacono S, Huang Y, Zhou JF, Duguay F, Chretien N, Welsh EA, Soares JW, Karatzas CN (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. Science 295:472–476

Leal-Egana A, Lang G, Mauerer C, Wickinghoff J, Weber M, Geimer S, Scheibel T (2012) Interactions of fibroblasts with different morphologies made of an engineered spider silk protein. Adv Eng Mater 14:B67–B75

Lee PA, Tullman-Ercek D, Georgiou G (2006) The bacterial twin-arginine translocation pathway. Annu Rev Microbiol 60:373–395

Lee H, Rho J, Messersmith PB (2009) Facile conjugation of biomolecules onto surfaces via mussel adhesive protein inspired coatings. *Adv Mater* 21:431

Lewis RV, Hinman M, Kothakota S, Fournier MJ (1996) Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. Protein Expr Purif 7:400–406

Li JB, Mohwald H, An ZH, Lu G (2005) Molecular assembly of biomimetic microcapsules. Soft Matter 1:259–264

Li LH, Puhl S, Meinel L, Germershaus O (2014) Silk fibroin layer-by-layer microcapsules for localized gene delivery. Biomaterials 35:7929–7939

Liebmann B, Huemmerich D, Scheibel T, Fehr M (2008) Formulation of poorly water-soluble substances using self-assembling spider silk protein. Colloids and Surfaces a-Physicochemical and Engineering Aspects 331:126–132

Lintz ES, Scheibel TR (2013) Dragline, egg stalk and byssus: a comparison of outstanding protein fibers and their potential for developing new materials. Adv Funct Mater 23:4467–4482

Liu Y, Shao Z, Vollrath F (2005) Relationships between supercontraction and mechanical properties of spider silk. Nat Mater 4:901–905

Lucas F, Rudall KM (1968) Extracellular fibrous proteins: the silks. In: Florkin M, Stotz EH (eds) Comprehensive biochemistry. Elsevier, NewYork

Luo J, Zhang LL, Peng QF, Sun MJ, Zhang YP, Shao HL, Hu XC (2014) Tough silk fibers prepared in air using a biomimetic microfluidic chip. Int J Biol Macromol 66:319–324

Maa YF, Hsu CC (1997) Feasibility of protein spray coating using a fluid-bed Wurster processor. Biotechnol Bioeng 53:560–566

Machado R, Da Costa A, Sencadas V, Garcia-Arevalo C, Costa CM, Padrao J, Gomes A, Lanceros-Mendez S, Rodriguez-Cabello JC, Casal M (2013) Electrospun silk-elastin-like fibre mats for tissue engineering applications. Biomed Mater 8

Madsen B, Shao ZZ, Vollrath F (1999) Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. Int J Biol Macromol 24:301–306

Maeda S, Kawai T, Obinata M, Fujiwara H, Horiuchi T, Saeki Y, Sato Y, Furusawa M (1985) Production of human alpha-interferon in silkworm using a baculovirus vector. Nature 315:592–594

Maitip J, Trueman HE, Kaehler BD, Huttley GA, Chantawannakul P, Sutherland TD (2015) Folding behavior of four silks of giant honey bee reflects the evolutionary conservation of aculeate silk proteins. Insect Biochem Mol Biol 59:72–79

Malda J, Visser J, Melchels FP, Jungst T, Hennink WE, Dhert WJA, Groll J, Hutmacher DW (2013) 25th anniversary article: engineering hydrogels for biofabrication. Adv Mater 25:5011–5028

Marolt D, Augst A, Freed LE, Vepari C, Fajardo R, Patel N, Gray M, Farley M, Kaplan D, Vunjak-Novakovic G (2006) Bone and cartilage tissue constructs grown using human bone marrow stromal cells, silk scaffolds and rotating bioreactors. Biomaterials 27:6138–6149

Marsh RE, Corey RB, Pauling L (1955) An investigation of the structure of silk fibroin. Biochim Biophys Acta 16:1–34

Megeed Z, Cappello J, Ghandehari H (2002) Genetically engineered silk-elastinlike protein polymers for controlled drug delivery. Adv Drug Deliv Rev 54:1075–1091

Megeed Z, Haider M, Li DQ, O'malley BW, Cappello J, Ghandehari H (2004) In vitro and in vivo evaluation of recombinant silk-elastinlike hydrogels for cancer gene therapy. J Control Release 94:433–445

Miao Y, Zhang Y, Nakagaki K, Zhao T, Zhao A, Meng Y, Nakagaki M, Park EY, Maenaka K (2005) Expression of spider flagelliform silk protein in *Bombyx mori* cell line by a novel Bac-to-Bac/BmNPV baculovirus expression system. Appl Microbiol Biotechnol 71:192–199

Mikos AG, Thorsen AJ, Czerwonka LA, Bao Y, Langer R, Winslow DN, Vacanti JP (1994) Preparation and characterization of poly(L-Lactic Acid) foams. Polymer 35:1068–1077

Moraes ML, Lima LR, Silva RR, Cavicchioli M, Ribeiro SJL (2013) Immunosensor based on immobilization of antigenic peptide NS5A-1 from HCV and silk fibroin in nanostructured films. Langmuir 29:3829–3834

Morse JC (1997) Phylogeny of trichoptera. Annu Rev Entomol 42:427–450

Nagarajan R, Drew C, Mello CM (2007) Polymer-micelle complex as an aid to electrospinning nanofibers from aqueous solutions. J Phys Chem C 111:16105–16108

Neuenfeldt M, Scheibel T (2014) Silks from insects: from natural diversity to applications. Insect molecular biology and ecology. CRC Press, Boca Raton

Ohshima Y, Suzuki Y (1977) Cloning of the silk fibroin gene and its flanking sequences. *Proc Natl Acad Sci U S A* 74:5363–5367

Omenetto FG, Kaplan DL (2010) New opportunities for an ancient material. Science 329:528–531

Osaki S (2012) Spider silk violin strings with a unique packing structure generate a soft and profound timbre. Phys Rev Lett 108:154301

Papov VV, Diamond TV, Biemann K, Waite JH (1995) Hydroxyarginine-containing polyphenolic proteins in the adhesive plaques of the marine mussel *Mytilus edulis*. J Biol Chem 270:20183–20192

Park WM, Champion JA (2014) Thermally triggered self-assembly of folded proteins into vesicles. J Am Chem Soc 136:17906–17909

Parker KD, Rudall KM (1957) The Silk of the egg-stalk of the green lace-wing fly: structure of the silk of chrysopa egg-stalks. Nature 179:905–906

Parkhe AD, Seeley SK, Gardner K, Thompson L, Lewis RV (1997) Structural studies of spider silk proteins in the fiber. J Mol Recognit : JMR 10:1–6

Pilotto F, Filosi M (1977) Relationship between collagen fibril diameters and body size – study of fish derma. Cell Tissue Res 182:119–131

Porter D, Guan J, Vollrath F (2013) Spider silk: super material or thin fibre? Adv Mater 25:1275–1279

Prince JT, Mcgrath KP, Digirolamo CM, Kaplan DL (1995) Construction, cloning, and expression of synthetic genes encoding spider dragline silk. Biochemistry 34:10879–10885

Qiu W, Huang Y, Teng W, Cohn CM, Cappello J, WU X (2010) Complete recombinant silk-elastinlike protein-based tissue scaffold. Biomacromolecules 11:3219–3227

Rabotyagova OS, Cebe P, Kaplan DL (2009) Self-assembly of genetically engineered spider silk block copolymers. Biomacromolecules 10:229–236

Rajkhowa R, Hu X, Tsuzuki T, Kaplan DL, Wang X (2012) Structure and biodegradation mechanism of milled *Bombyx mori* silk particles. Biomacromolecules 13:2503–2512

Rammensee S, Slotta U, Scheibel T, Bausch AR (2008) Assembly mechanism of recombinant spider silk proteins. Proc Natl Acad Sci U S A 105:6590–6595

Raphel J, Parisi-Amon A, Heilshorn SC (2012) Photoreactive elastin-like proteins for use as versatile bioactive materials and surface coatings. J Mater Chem 22:19429–19437

Reneker DH, Yarin AL (2008) Electrospinning jets and polymer nanofibers. Polymer 49:2387–2425

Riekel C, Vollrath F (2001) Spider silk fibre extrusion: combined wide- and small-angle X-ray microdiffraction experiments. Int J Biol Macromol 29:203–210

Rising A, Hjalm G, Engstrom W, Johansson J (2006) N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. Biomacromolecules 7:3120–3124

Römer L, Scheibel T (2008) The elaborate structure of spider silk: structure and function of a natural high performance fiber. Prion 2:154–161

Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol 5:172

Rudall KM, Kenchington W (1971) Arthropod silks: the problem of fibrous proteins in animal tissues. Annu Rev Entomol 16:73–96

Santoso S, Hwang W, Hartman H, Zhang SG (2002) Self-assembly of surfactant-like peptides with variable glycine tails to form nanotubes and nanovesicles. Nano Lett 2:687–691

Schacht K, Scheibel T (2011) Controlled hydrogel formation of a recombinant spider silk protein. Biomacromolecules 12:2488–2495

Schacht K, Jungst T, Schweinlin M, Ewald A, Groll J, Scheibel T (2015) Biofabrication of cell-loaded 3D spider silk constructs. Angewandte Chemie-International Edition 54:2816–2820

Schacht K, Vogt J, Scheibel T (2016). Foams made of engineered recombinant spider silk proteins as 3D scaffolds for cell growth. ACS Biomater Sci Eng, dx.doi.org/10.1021/acsbiomaterials.5b00483

Scheibel T (2004) Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. Microb Cell Factories 3:14

Scheller J, Conrad U (2005) Plant-based material, protein and biodegradable plastic. Curr Opin Plant Biol 8:188–196

Scheller J, Guhrs KH, Grosse F, Conrad U (2001) Production of spider silk proteins in tobacco and potato. Nat Biotechnol 19:573–577

Scheller J, Henggeler D, Viviani A, Conrad U (2004) Purification of spider silk-elastin from transgenic plants and application for human chondrocyte proliferation. Transgenic Res 13:51–57

Sehnal F, Akai H (1990) Insect silk glands: their types, development and function, and effects of environmental factors and morphogenetic hormones on them. Int J Insect Morphol Embryol 19:79–132

Seidel A, Liivak O, Calve S, Adaska J, Ji GD, Yang ZT, Grubb D, Zax DB, Jelinski LW (2000) Regenerated spider silk: processing, properties, and structure. Macromolecules 33:775–780

Shchepelina O, Drachuk I, Gupta MK, Lin J, Tsukruk VV (2011) Silk-on-silk layer-by-layer microcapsules. *Adv Mater* 23:4655

Shi J, Lua S, Du N, Liu X, Song J (2008) Identification, recombinant production and structural characterization of four silk proteins from the Asiatic honeybee *Apis cerana*. Biomaterials 29:2820–2828

Silva-Zacarin EC, De Moraes R L S, Taboga SR (2003) Silk formation mechanisms in the larval salivary glands of *Apis mellifera* (Hymenoptera: Apidae). J Biosci 28:753–764

Slotta U, Hess S, Spiess K, Stromer T, Serpell L, Scheibel T (2007) Spider silk and amyloid fibrils: a structural comparison. Macromol Biosci 7:183–188

Slotta UK, Rammensee S, Gorb S, Scheibel T (2008) An engineered spider silk protein forms microspheres. Angew Chem Int Ed 47:4592–4594

Spiess K, Lammel A, Scheibel T (2010) Recombinant spider silk proteins for applications in biomaterials. Macromol Biosci 10:998–1007

Spiess K, Ene R, Keenan CD, Senker J, Kremer F, Scheibel T (2011) Impact of initial solvent on thermal stability and mechanical properties of recombinant spider silk films. J Mater Chem 21:13594–13604

Sponner A, Schlott B, Vollrath F, Unger E, Grosse F, Weisshart K (2005) Characterization of the protein components of *Nephila clavipes* dragline silk. Biochemistry 44:4727–4736

Stephens JS, Fahnestock SR, Farmer RS, Kiick KL, Chase DB, Rabolt JF (2005) Effects of electrospinning and solution casting protocols on the secondary structure of a genetically engineered dragline spider silk analogue investigated via fourier transform Raman spectroscopy. Biomacromolecules 6:1405–1413

Stewart RJ (2011) Protein-based underwater adhesives and the prospects for their biotechnological production. Appl Microbiol Biotechnol 89:27–33

Stewart RJ, Wang CS (2010) Adaptation of caddisfly larval silks to aquatic habitats by phosphorylation of h-fibroin serines. Biomacromolecules 11:969–974

Sun W, Yu H, Shen Y, Banno Y, Xiang Z, Zhang Z (2012) Phylogeny and evolutionary history of the silkworm. Sci China Life Sci 55:483–496

Sundaray B, Subramanian V, Natarajan TS, Xiang RZ, Chang CC, Fann WS (2004) Electrospinning of continuous aligned polymer fibers. Appl Phys Lett 84:1222–1224

Sutherland TD, Campbell PM, Weisman S, Trueman HE, Sriskantha A, Wanjura WJ, Haritos VS (2006) A highly divergent gene cluster in honey bees encodes a novel silk family. Genome Res 16:1414–1421

Sutherland TD, Weisman S, Trueman HE, Sriskantha A, Trueman JW, Haritos VS (2007) Conservation of essential design features in coiled coil silks. Mol Biol Evol 24:2424–2432

Sutherland TD, Young JH, Weisman S, Hayashi CY, Merritt DJ (2010) Insect silk: one name, many materials. Annu Rev Entomol 55:171–188

Sutherland TD, Weisman S, Walker AA, Mudie ST (2012) The coiled coil silk of bees, ants, and hornets. Biopolymers 97:446–454

Sutherland TD, Trueman HE, Walker AA, Weisman S, Campbell PM, Dong Z, Huson MG, Woodhead AL, Church JS (2014) Convergently-evolved structural anomalies in the coiled coil domains of insect silk proteins. J Struct Biol 186:402–411

Tamada Y (2005) New process to form a silk fibroin porous 3-D structure. Biomacromolecules 6:3100–3106

Tamura T, Thibert C, Royer C, Kanda T, Abraham E, Kamba M, Komoto N, Thomas JL, Mauchamp B, Chavancy G, Shirk P, Fraser M, Prudhomme JC, Couble P (2000) Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. Nat Biotechnol 18:81–84

Tanaka K, Mori K, Mizuno S (1993) Immunological identification of the major disulfide-linked light component of silk fibroin. J Biochem 114:1–4

Tanaka K, Inoue S, Mizuno S (1999a) Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by *Bombyx mori*. Insect Biochem Mol Biol 29:269–276

Tanaka K, Kajiyama N, Ishikura K, Waga S, Kikuchi A, Ohtomo K, Takagi T, Mizuno S (1999b) Determination of the site of disulfide linkage between heavy and light chains of silk fibroin produced by *Bombyx mori*. Biochim Biophys Acta 1432:92–103

Teng W, Huang Y, Cappello J, Wu X (2011) Optically transparent recombinant silk-elastinlike protein polymer films. J Phys Chem B 115:1608–1615

Tokareva O, Michalczechen-Lacerda VA, Rech EL, Kaplan DL (2013) Recombinant DNA production of spider silk proteins. Microb Biotechnol 6:651–663

Tomita M (2011) Transgenic silkworms that weave recombinant proteins into silk cocoons. Biotechnol Lett 33:645–654

Tsukada M, Khan MM, Inoue E, Kimura G, Hun JY, Mishima M, Hirabayashi K (2010) Physical properties and structure of aquatic silk fiber from *Stenopsyche marmorata*. Int J Biol Macromol 46:54–58

Tucker CL, Jones JA, Bringhurst HN, Copeland CG, Addison JB, Weber WS, Mou Q, Yarger JL, Lewis RV (2014) Mechanical and physical properties of recombinant spider silk films using organic and aqueous solvents. Biomacromolecules 15:3158–3170

Van Beek JD, Hess S, Vollrath F, Meier BH (2002) The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. *Proc Natl Acad Sci U S A* 99:10266–10271

Vasconcelos A, Freddi G, Cavaco-Paulo A (2008) Biodegradable materials based on silk fibroin and keratin. Biomacromolecules 9:1299–1305

Vauthey S, Santoso S, Gong HY, Watson N, Zhang SG (2002) Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. *Proc Natl Acad Sci U S A* 99:5355–5360

Vendrely C, Scheibel T (2007) Biotechnological production of spider-silk proteins enables new applications. Macromol Biosci 7:401–409

Vepari C, Kaplan DL (2007) Silk as a biomaterial. Prog Polym Sci 32:991–1007

Vollrath F (2000) Strength and structure of spiders silks. J Biotechnol 74:67–83

Vollrath F (2006) Spider silk: thousands of nano-filaments and dollops of sticky glue. Curr Biol : CB 16:R925–R927

Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. Nature 410:541–548

Vollrath F, Porter D (2006) Spider silk as archetypal protein elastomer. Soft Matter 2:377–385

Wang YZ, Kim UJ, Blasioli DJ, Kim HJ, Kaplan DL (2005) In vitro cartilage tissue engineering with 3D porous aqueous-derived silk scaffolds and mesenchymal stem cells. Biomaterials 26:7082–7094

Wang Y, Sanai K, Wen H, Zhao T, Nakagaki M (2010) Characterization of unique heavy chain fibroin filaments spun underwater by the caddisfly *Stenopsyche marmorata* (Trichoptera; Stenopsychidae). Mol Biol Rep 37:2885–2892

Wang CS, Ashton NN, Weiss RB, Stewart RJ (2014) Peroxinectin catalyzed dityrosine crosslinking in the adhesive underwater silk of a casemaker caddisfly larvae, Hysperophylax occidentalis. Insect Biochem Mol Biol 54:69–79

Weichert N, Hauptmann V, Helmold C, Conrad U (2016) Seed-specific expression of spider silk protein multimers causes long-term stability. Front Plant Sci 7:6

Weisman S, Trueman HE, Mudie ST, Church JS, Sutherland TD, Haritos VS (2008) An unlikely silk: the composite material of green lacewing cocoons. Biomacromolecules 9:3065–3069

Weisman S, Okada S, Mudie ST, Huson MG, Trueman HE, Sriskantha A, Haritos VS, Sutherland TD (2009) Fifty years later: the sequence, structure and function of lacewing cross-beta silk. J Struct Biol 168:467–475

Weisman S, Haritos VS, Church JS, Huson MG, Mudie ST, Rodgers AJ, Dumsday GJ, Sutherland TD (2010) Honeybee silk: recombinant protein production, assembly and fiber spinning. Biomaterials 31:2695–2700

Wen H, Lan X, Zhang Y, Zhao T, Wang Y, Kajiura Z, Nakagaki M (2010) Transgenic silkworms (*Bombyx mori*) produce recombinant spider dragline silk in cocoons. Mol Biol Rep 37:1815–1821

Widmaier DM, Tullman-Ercek D, Mirsky EA, Hill R, Govindarajan S, Minshull J, Voigt CA (2009) Engineering the Salmonella type III secretion system to export spider silk monomers. Mol Syst Biol 5:309

Wiggins GB (2004) Caddisflies: the underwater architects. University of Toronto Press, Toronto

Wittmer CR, Hu X, Gauthier PC, Weisman S, Kaplan DL, Sutherland TD (2011) Production, structure and in vitro degradation of electrospun honeybee silk nanofibers. Acta Biomater 7:3789–3795

Wohlrab S, Spieß K, Scheibel T (2012) Varying surface hydrophobicities of coatings made of recombinant spider silk proteins. J Mater Chem 22:22050–22054

Woolfson DN (2005) The design of coiled-coil structures and assemblies. Adv Protein Chem 70:79–112

Xia XX, Qian ZG, Ki CS, Park YH, Kaplan DL, Lee SY (2010) Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. Proc Natl Acad Sci USA 107:14059–14063

Xu H (2014) The advances and perspectives of recombinant protein production in the silk gland of silkworm *Bombyx mori*. Transgenic Res 23:697–706

Xu HT, Fan BL, Yu SY, Huang YH, Zhao ZH, Lian ZX, DAI YP, Wang LL, Liu ZL, Fei J, Li N (2007) Construct synthetic gene encoding artificial spider dragline silk protein and its expression in milk of transgenic mice. Anim Biotechnol 18:1–12

Xu L, Tremblay ML, Orrell KE, Leclerc J, Meng Q, Liu XQ, Rainey JK (2013) Nanoparticle self-assembly by a highly stable recombinant spider wrapping silk protein subunit. FEBS Lett 587:3273–3280

Yamaguchi K, Kikuchi Y, Takagi T, Kikuchi A, Oyama F, Shimura K, Mizuno S (1989) Primary structure of the silk fibroin light chain determined by cDNA sequencing and peptide analysis. J Mol Biol 210:127–139

Yanagisawa S, Zhu Z, Kobayashi I, Uchino K, Tamada Y, Tamura T, Asakura T (2007) Improving cell-adhesive properties of recombinant *Bombyx mori* silk by incorporation of collagen or fibronectin derived peptides produced by transgenic silkworms. Biomacromolecules 8:3487–3492

Yang M, Tanaka C, Yamauchi K, Ohgo K, Kurokawa M, Asakura T (2008) Silklike materials constructed from sequences of *Bombyx mori* silk fibroin, fibronectin, and elastin. J Biomed Mater Res A 84:353–363

Yao J, Yanagisawa S, asakura T (2004) Design, expression and characterization of collagen-like proteins based on the cell adhesive and crosslinking sequences derived from native collagens. J Biochem 136:643–649

Yonemura N, Sehnal F, Mita K, Tamura T (2006) Protein composition of silk filaments spun under water by caddisfly larvae. Biomacromolecules 7:3370–3378

Young SL, Gupta M, Hanske C, Fery A, Scheibel T, Tsukruk VV (2012) Utilizing conformational changes for patterning thin films of recombinant spider silk proteins. Biomacromolecules 13:3189–3199

Yucel T, Cebe P, Kaplan DL (2009) Vortex-induced injectable silk fibroin hydrogels. Biophys J 97:2044–2050

Zeplin PH, Berninger AK, Maksimovikj NC, Van Gelder P, Scheibel T, Walles H (2014a) Improving the biocompatibility of silicone implants using spider silk coatings: immunohistochemical analysis of capsule formation. Handchir Mikrochir Plast Chir 46:336–341

Zeplin PH, Maksimovikj NC, Jordan MC, Nickel J, Lang G, Leimer AH, Roemer L, Scheibel T (2014b) Spider silk coatings as a bioshield to reduce periprosthetic fibrous capsule formation. Adv Funct Mater 24:2658–2666

Zhang Y, Hu J, Miao Y, Zhao A, Zhao T, Wu D, Liang L, Miikura A, Shiomi K, Kajiura Z, Nakagaki M (2008) Expression of EGFP-spider dragline silk fusion protein in BmN cells and larvae of silkworm showed the solubility is primary limit for dragline proteins yield. Mol Biol Rep 35:329–335

Zhang K, Duan H, Karihaloo BL, Wang J (2010) Hierarchical, multilayered cell walls reinforced by recycled silk cocoons enhance the structural integrity of honeybee combs. Proc Natl Acad Sci U S A 107:9502–9506

Zhao QH, Li BY (2008) pH-controlled drug loading and release from biodegradable microcapsules. Nanomed Nanotechnol Biol Med 4:302–310

Zhou CZ, Confalonieri F, Jacquet M, Perasso R, LI ZG, Janin J (2001) Silk fibroin: structural implications of a remarkable amino acid sequence. Proteins 44:119–122

Zhu JM, Marchant RE (2011) Design properties of hydrogel tissue-engineering scaffolds. Expert Rev Med Devices 8:607–626

Zhu B, Wang H, Leow WR, Cai Y, Loh XJ, Han MY, Chen X (2015) Silk fibroin for flexible electronic devices. Adv Mater 28(22):4250–4265

Chapter 17
Biomaterials Made from Coiled-Coil Peptides

Vincent Conticello, Spencer Hughes, and Charles Modlin

Contents

Abstract The development of biomaterials designed for specific applications is an important objective in personalized medicine. While the breadth and prominence of biomaterials have increased exponentially over the past decades, critical challenges remain to be addressed, particularly in the development of biomaterials that exhibit highly specific functions. These functional properties are often encoded within the molecular structure of the component molecules. Proteins, as a consequence of their structural specificity, represent useful substrates for the construction of functional biomaterials through rational design. This chapter provides an in-depth survey of biomaterials constructed from coiled-coils, one of the best-understood protein structural motifs. We discuss the utility of this structurally diverse and functionally tunable class of proteins for the creation of novel biomaterials. This discussion illustrates the progress that has been made in the development of coiled-coil biomaterials by showcasing studies that bridge the gap between the academic science and potential technological impact.

Keywords Coiled-coils • Biomaterials • Hydrogels • Nanotubes • Nanomedicine

V. Conticello (✉) • S. Hughes • C. Modlin
Emory University Chemistry Department, 1515 Dickey Drive, Atlanta 30322, Georgia
e-mail: vcontic@emory.edu

Abbreviations

PHPMA	poly (N-(2-hydroxypropyl))methylacrylamide
MTX	methotrexate
RHCC	Right-handed coiled-coil
EPR	Enhanced permeability and retention
NHL	Non-Hodgkin's Lymphoma
RGD	Arginine-Glycine-Aspartate
LZ	leucine zipper
HPMA	N(2-hydroxypropyl)methylacrylamide
DAMA	N',N'-(dicarboxymethylaminopropyl)methacrylamide
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium
COMP	cartilage oligomeric matrix protein
PEGDA	polyethylene glycol diacrylate
hSAFs	hydrogelating self-assembling fibres
SAFs	self-assembling fibres
AUC	analytical ultra centrifugation
SARS	Severe Acute Respiratory Syndrome
SMCC	The N-hydroxysuccinimide ester of 4-(N-maleimidomethyl) cyclohexanecarboxylic acid

17.1 Introduction: The Emerging Role of Biomaterials

Biomaterials derived from designed peptide and protein sequences have become the focus of significant scientific research effort over the last two decades. Rapid developments in genome sequencing, bioinformatic analysis, atomic resolution structural determination, and computational protein design, in combination with extant synthetic and analytical methods, present an unprecedented opportunity to create peptide- and protein-based biomaterials of defined structure and controllable function. Progress toward the goal of creating tailorable biomaterials for applications in nanotechnology and human health has been impressive, although the practical promise has yet to be realized on a substantial scale. One strategy has involved the identification of stable structural motifs that can serve as flexible scaffolds for the introduction of function. This approach requires the identification of protein folds that can accommodate substantial sequence modification without compromising structure or function. Ideally, these protein motifs would be highly designable and amenable to computational structural prediction. The α-helical coiled-coil motif is the most studied protein fold and fulfills many of the criteria that one would envision as a scaffold for biomaterials design.

In recent years, the well-known coiled-coil motif has been widely employed in the generation of novel biomaterials. The canonical coiled-coil, comprised of two

(or more) right-handed α-helices, is a left-handed super-helix wherein the constituent α-helices have a crossing angle (angle between two helices of a coiled-coil (Lupas 1996) of roughly 20° (Lupas and Gruber 2005)). However, minute mutations to the primary sequence can produce profound changes that can cause the coiled-coil to deviate from the classical coiled-coil definition referenced above (Zimenkov et al. 2006; Wood et al. 2014). Such deviations, such as stutters and stammers, often alter the resultant biophysical properties of the coiled-coil (Gruber and Lupas 2003). Coiled-coils present multiple advantages as biomaterials scaffolds, including:

1. Well-understood design rules (Lupas and Gruber 2005)
2. Accessibility of a variety of oligomeric states (Harbury et al. 1993)
3. Dynamic responsiveness to environmental factors (Zimenkov et al. 2006; Dublin and Conticello 2008; Banwell et al. 2009)
4. Versatility in sequence and structure
5. Biocompatibility

As discussed in previous chapters, coiled-coils can be designed *de novo* using a handful of simple rules that can be implemented computationally to select for oligomerization state with a great degree of reliability (Thompson et al. 2014; Huang et al. 2005; Wood et al. 2014).

Hydrophobic residues usually occupy a majority of the core *a* and *d* positions of the heptad repeats. Charge complementarity can be employed at proximal positions to control helix alignment or promote the formation of heteromeric assemblies. Sequence control of the residues that occupy the (*a, d, e, g*) positions, which define the extended helix-helix interface, can be employed to control the oligomerization state in the range from 2 to 7 helices. In contrast, design rules (Grigoryan and DeGrado 2011) to specify the formation of specific super-secondary or tertiary structures are not as readily apparent for β-strand-based biomaterials, or, at this juncture, more complex structural motifs. By explicit design, coiled-coils have been developed with environmental "switches"; in which a pH drop might cause dissociation of a coiled-coil into its constituent helices, or a metal ion may induce a registry shift (Fig. 17.1), leading to an exclusive population of one oligomeric structure (Anzini et al. 2013).

As evidenced by their ubiquity in the proteomes of sequenced organisms, coiled-coil proteins comprise varied sequences, adopt a variety of native structures, and display a range of functions (Testa et al. 2009). The scope of coiled-coil sequences and structures has been further expanded through the carefully applied use of rational design. Non-canonical amino acids can be incorporated at specified sites in the sequences using either solid-phase peptide synthesis methods or supplementation of growth media used in *in vitro* or *in vivo* protein expression. The introduction of non-native chemical functionality can further expand the material properties of this flexible scaffold. Indeed, coiled-coils are a versatile and robust scaffold for generating a diverse range of functional biomaterials that can be tailored for specific applications through modification of the sequence at structurally permissive sites.
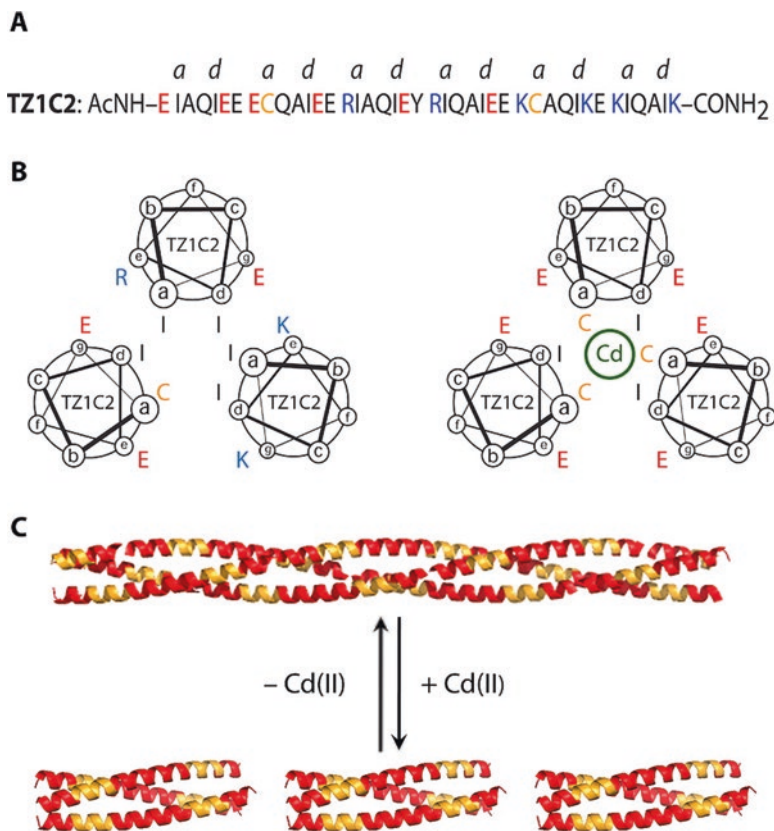
Fig. 17.1 Design of a metal-ion induced registry shift into a trimeric coiled-coil structure. Briefly, Cd^{2+} ions coordinate with cysteine sidechains in the coiled-coil's hydrophobic core, locking them into a trimeric state, rather than the off-register fibrillar structure seen in non-metallated samples. (**a**) Single-letter amino acid sequence of the TZ1C2 peptide, with the *a* and *d* positions labelled to highlight the hydrophobic core of the coiled-coil structure. (**b**) Helical-wheel representations of the out-of-register (*left*), metal-free peptide, and the in-register (*right*), metal-coordinated peptide. (**c**) Fibre representation of the metal-ion oligomerization switch (Reprinted with permission from Anzini et al. (2013) Copyright 2013 *J Am Chem Soc*)

17.2 Coiled-Coil Therapeutics

A major problem facing modern medicine is the dearth of strategies for treating diseases that present differently in each patient, e.g. cancer. The need for easily customizable treatments has led to the development of a broad field known as nano-theranostics (a portmanteau derived from therapeutics and diagnostics), which seeks to use nano-scale materials to image and/or treat diseases. A dominant strategy in nano-theranostics is to specifically target diseased cells for the delivery of small doses of extremely cytotoxic low molecular weight drugs. A clear challenge in this field is thus developing site-specific drug delivery to reduce side effects from

apoptosis of healthy cells. Coiled-coil motifs have recently been explored as potential nano-theranostic agents, due to their lack of immunogenicity, exquisite sensitivity to biologically relevant environmental changes (such as pH), and well-known engineering rules. This inherent designability makes coiled-coils an attractive scaffold from which to develop materials for nano-scale treatments.

Coiled-coil-based therapeutics can be broken up into two main categories: coiled-coil drug carriers, and drug-free coiled-coil systems. The first of these categories is more familiar to those studying nano-medicine. Essentially, a low molecular weight drug is encapsulated by a self-assembling coiled-coil, which dissociates upon some environmental change (typically pH, which can vary in a diseased state) (Yu 2002). The low molecular weight drug is generally highly cytotoxic and, as such, the design focus for this approach is towards target specificity (Pola et al. 2013). These systems also offer the potential for multi-valent drug binding, which improves the loading efficiency by minimizing the entropic penalty for binding. Assuming one chain in a heterodimeric coiled-coil system contains a single binding site, its sequence may be doubled to introduce a second binding site (Assal et al. 2015). Recently, Klok and co-workers investigated a heterodimeric coiled-coil carrier comprising two peptides (K_3 and E_3), which self-assemble at pH 7.0, but dissociate at pH 5.0 (endosomal pH), thereby releasing their cargo (Apostolovic et al. 2010, 2011). These peptides were bound to a poly (N-(2-hydroxypropyl)-methacrylamide (PHPMA) polymer backbone and designed to carry an anti-cancer drug, methotrexate (MTX) in one case, and a fluorescent dye in another. By combining these two technologies, it follows that one could monitor cellular uptake of the drug carrier while simultaneously treating the patient. Though uptake of the drug into the endosomes was shown, the distribution of the drug throughout the organelles has not yet been obtained.

In a more direct therapeutic solution, Thanasupawat et al., recently reported a coiled-coil-based system with improved selectivity that targets gliomas (malignant tumours that bud from glial cells in either the brain or the spine) (Thanasupawat et al. 2015). Briefly, they developed a right-handed coiled-coil (RHCC) comprising four identical alpha helices. Each alpha helix presents a methionine residue on the solvent-exposed surface of the tube, through which it coordinates a Platinum (IV) ion derived from tetrachloroplatinate (Fig. 17.2). The complexed Pt(IV) displayed a stronger selectivity for the target tumour and less cytotoxicity than the more labile cis-platin derivative of the RHCC. The RHCC essentially serves as a carrier for four equivalents of the Pt(IV) complex, which acts as a pro-drug that is activated via conversion to the therapeutic Pt(II) state in the reductive environment of the tumour cell. This phenomenon affords higher efficacy and fewer side effects, resulting in minimization of treatment cost and duration.

The second approach involves grafting of coiled-coil peptides to a synthetic polymer backbone; upon self-assembly, the polymer-peptide complex induces a physiological change in the cell that triggers apoptosis (Pechar et al. 2011). Although this approach has only recently emerged, it presents a strong advantage over drug-based therapies; namely, none of the components acting alone can induce apoptosis. Assuming specific bio-recognition of the constituent peptides, apoptosis will only

Fig. 17.2 Complex comprising four alpha-helices in a right-handed coiled coil conformation with four coordinated Pt(IV) ions (*red spheres*) (Reprinted with permission from Thanasupawat et al. (2015) Copyright 2015 *Nanomedicine*)
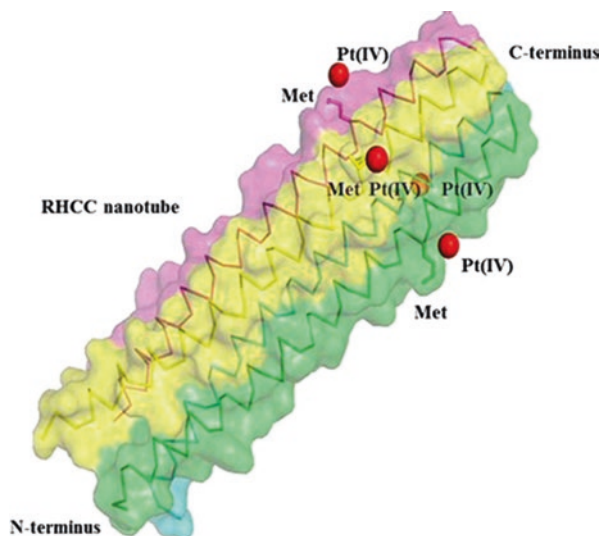
occur at the diseased cell. This represents a new paradigm in nanomedicine, and was pioneered by Kopeček and co-workers (Wu et al. 2012; Chu and Kopeček 2015). These researchers developed the idea of attaching high molecular weight polymers to small coiled-coil-forming peptides. The biocompatible polymer is linked to one peptide to increase endocytosis by cancer cells, via the enhanced permeability and retention (EPR) effect. This effect states that due to the increased vasculature and poor draining of tumour cells, high molecular weight polymers are more likely to be endocytosed than smaller species, and are also less likely to be drained through the lymphocytes. A second peptide, complementary to the first, is conjugated to a receptor recognition motif, which directs the peptide to a cell type of interest. The Kopeček group explored the use of these materials for the treatment of Non-Hodgkin's Lymphoma (NHL). In NHL, a specific surface receptor (CD20) is present on >95 % of tumourous cells. CD20 is also present on healthy B cells, but not on stem cells. It has been shown that the temporary B cell depletion is not deleterious. The recognition motif present on a specific antibody (Fab') is used as the targeting moiety in this model system. Fab' and its peptide partner bind to the CD20 receptors, and the peptide-polymer conjugates are effectively drawn towards the B cells. At this point, the peptides interact to form a heterodimeric coiled-coil, which induces cross-linking of the CD20 receptors. The cross-linking of CD20 receptors results in apoptosis, and effective remediation of the disease (Fig. 17.3).

17.3 Coiled-Coil Hydrogels

Hydrogels are formed from covalent or non-covalent cross-linking of individual polymer chains into a three-dimensional water-swollen network. Hydrogel materials offer a wealth of potential applications (Elisseeff 2008), in that the biological,
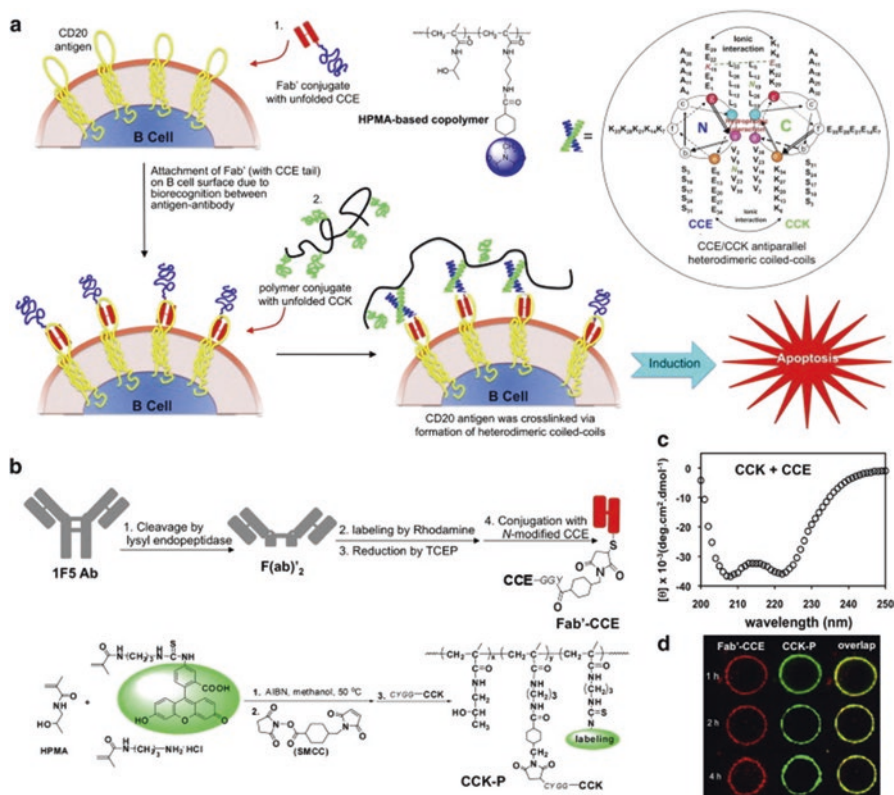
Fig. 17.3 Schematic representation of the drug-free coiled-coil-based therapeutics employed by the Kopeček group. The unfolded CCE peptide (*blue coil*) is chemically conjugated to the Fab' recognition domain (*red*), which directs it to the CD20 antigen (*yellow*). Unfolded CCK peptide (*green coil*) is directed to the cell surface via the EPR effect and conjugation to the HPMA copolymer backbone (*black coil*). Coiled-coil formation by the two peptides results in CD20 antigen crosslinking and subsequent apoptosis (Reprinted with permission from Wu et al. (2012) Copyright 2012 *J Control Release*)

chemical, and mechanical properties may be tailored for particular applications through modification of the structure of the polymer and cross-linker. They have innate capacity to absorb and retain large volumes of water relative to the mass of the hydrogel itself (Elisseeff 2008; Xu and Kopeček 2008; Yao et al. 2014). Moreover, hydrogels are often highly biocompatible due to their capacity for water absorption, physical elasticity, and relative softness (Kopeček 2007, Yao et al. 2014). The aforementioned qualities make hydrogels a convincing substitute for native extracellular matrix (Slaughter et al. 2009; Yao et al. 2014), as well as suitable candidates for cell culture applications (Frampton et al. 2011; Jung et al. 2011), biomedical uses such as wound treatment (Wong et al. 2011; Lin et al. 2011), and drug delivery (Garbern et al. 2011; Guziewicz et al. 2011). Unsurprisingly, hydrogels were the first biomaterials developed for medical uses in humans (Kopeček

2007; Kopeček and Yang 2012). As biomedical materials, coiled-coil-based hydrogels offer an alternative to the more commonly employed β-sheet hydrogels. One potential concern with the latter materials is that the β-sheet morphology is analogous to amyloid structures, such as those observed in prion and amyloid diseases. These toxic assemblies often develop via a misfolding of endogenous proteins and can be auto-catalytically propagated in the presence of a β-sheet template (Fletcher et al. 2011). Due to the risk of initiating protein misfolding disease states, implementation of β-sheet-based hydrogels must be carefully scrutinized for the manifestation of deleterious physiological effects (Fletcher et al. 2011). Biomaterials based on coiled-coil structures do not display similar behaviour, as self-assembly in most cases is restricted to the formation of closed oligomers. This section will cover the past decade of advances within coiled-coiled-based hydrogels.

Hydrogels based on coiled-coil self-association can offer several notable advantages (Kopeček 2007). The coiled-coil is an exceptionally well understood structural motif, when compared to other protein folding domains (*vide supra*) (Lupas and Gruber 2005). Due to the wealth of information regarding the coiled-coil's structure and properties (in comparison to other super-secondary and tertiary structures), the coiled-coil has the potential to serve as an ideal 'tuning knob' for hydrogel properties. Though the coiled-coil is certainly not the only factor that governs hydrogel properties, it represents the most facile mechanistic handle. Tuning coiled-coil properties can involve a procedure as simple as altering the sequence of amino acids within the constituent alpha helices, allowing the researcher to alter hydrogel properties with relative ease and efficiency. Coiled-coils offer versatility in terms of oligomeric state as well (Harbury et al. 1993). Moreover, due to the widespread prevalence of coiled-coils within native proteins, coiled-coil-based biorecognition domains are also pervasive throughout the natural world. Co-opting such domains may yield significant advantages in the engineering and design of hydrogels, as doing so may offer a conduit through which artificially engineered hydrogels may interact with natural systems with high specificity (Kopeček 2007).

For the purpose of this review two primary categories for *self-assembling* hydrogels exist: those that consist only of peptide/protein and those that are based on hybrid materials (Kopeček 2007). The latter incorporate non-peptidic moieties within the component molecules of the hydrogel. The most common protein hydrogels derived from coiled-coils are based on an ABA triblock copolymer architecture. Petka et al. developed the first ABA triblock polymer system for the fabrication of hydrogels (Petka et al. 1998). Subsequently, Xu and Kopeček further developed these materials. These *ABA* triblock copolymers comprise a sequence in which coiled-coil motifs occupy the endblock domains (*A*) while a hydrophilic random coil sequence (*B*) defines the midblock (Fig. 17.4). Self-association of alpha helical (*A*)-end-blocks initiates cross-linking of the polymers into a water-swollen network in which the hydrophilic (*B*) mid-block connects the coiled-coil domains (Kopeček 2007).

The self-assembly of these ABA block copolymers is driven by two interconnected forces: hydrophobic collapse of A ends and polar association of water with the B interior region. One advantage of hydrogel formation from these ABA block
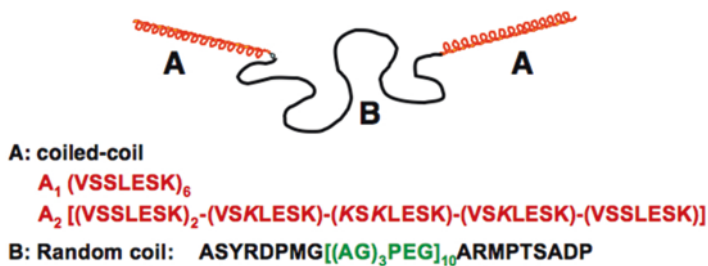
Fig. 17.4 Graphical representation of ABA triblock copolymer system pioneered by the Kopeček group, wherein *A* ends (shown in *red*) are alpha helices and the *B* middle (shown in *black*) is a random coil. The *A* helix ends associate with neighbouring *A* helices to form coiled-coils, generating a coiled-coil-based hydrogel. The sequences used by the Kopeček group for *A* and *B* portions are listed below, though there exists inherent mutability to these sequences, allowing for potential customization of a triblock hydrogel's biophysical properties in downstream applications (Reprinted with permission from Kopeček et al. (2007) Copyright 2007 *Biomaterials*)

copolymers is their reversibility, wherein virtual cross-links resultant from the coiled-coil interactions can be reversibly denatured, yielding a gel to sol transition. However, the A blocks retain the capacity to re-form the coiled-coil structure upon removal of the denaturing conditions. Moreover, the interaction between helices is highly tunable; minute changes in sequence, especially to the alpha helical regions can produce discernible effects on the biophysical properties of the resultant morphologies (Kopeček 2007). Xu and Kopeček noted that the physical characteristics of their ABA triblock and AB diblock hydrogels depended on the relative lengths of the A block and B block, the sequence composition, and the introduction of specific electrostatic interactions. Tirrell and co-workers programmed further specificity into these systems through diversification of the polypeptide sequence to create an ABC triblock system. In this architecture, the A and C domains comprise different coiled-coil sequences that exhibit little tendency to associate heteromerically, that is, the A and C blocks are structurally orthogonal (Shen et al. 2006). The presence of these orthogonal coiled-coil sequences inhibits the formation of intramolecular loops, which were prevalent within hydrogels derived from the ABA triblock systems, and, consequently, lowers the rate of erosion of free polymer from the surface of the resultant hydrogel.

An additional advantage of the coiled-coil motif lies in its ability to accommodate mutations at specific positions while retaining its structural integrity (Harbury et al. 1993, 1994). As a consequence, sequences that encode biological functions can be introduced into the scaffold of the coiled-coil structural units. This structural robustness enabled Huang et al. to implant the RGD domain between two hexaheptad sequences (A2) derived from a previously published leucine zipper (LZ) coiled-coil peptide. The RGD domain was separated from each A2 sequence by a nine amino acid spacer (C1). A *C*-terminal tail (C2), containing cysteine, was also added, completing an RGD-containing system constructed as follows: A2-C1-

RGDS-C1-A2-C2. The insertion of RGDS imbued cell adhesion to the hydrogel (Huang et al. 2014). Specifically, Huang et al. found that RGD sequences introduced into LZ hydrogels promoted the attachment (as determined via MTS cell proliferation assay kit, a colorimetric analysis which measures the reduction of 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium, or MTS) and proliferation of metabolically active cells at quantitatively higher levels than the control system: A2-C1-C1-A2-C2, which lacked the RGD domain. These results were corroborated *in vivo* using human marrow stem cells in a murine host system, which demonstrated cell attachment, growth and even neovascularity, thus proving that a mutable, tunable, coiled-coil-based hydrogel has potential for tissue engineering applications (Huang et al. 2014). This "customization" of a hydrogel is possible primarily due to the structural stability encoded within the coiled-coil LZ domain, and its ability to incorporate mutations near it without loss of its structure/function relationship.

Hybrid hydrogels are derived from self-assembly of synthetic polymers to which coiled-coil sequences have been covalently attached, usually at the termini, to afford ABA-type protein-polymer conjugates (Kopeček 2007). These hybrid materials combine the specificity of coiled-coil interactions with the well-known mechanical properties and biocompatibility of synthetic biomaterials. Kopecek and Yang (2012) linked N-(2-Hydroxypropyl) methacrylamide (HPMA) to (N',N'-dicarboxymethylaminopropyl) methacrylamide (DAMA) to create a copolymer that served as the main polymer component of the hydrogel. The HPMA polymer, a hydrophilic molecule, promotes the water absorption and swelling of the hydrogel. DAMA, a metal-chelating agent, provided a mechanism to couple the polymer chain to the peptide moiety of the hydrogel via Ni^{2+} chelation between DAMA and a His-tag domain within the peptide. The HPMA-DAMA copolymer was conjugated to two differing coiled-coil-forming peptide sequences with differing lengths and amino acid compositions: CC1 and CC2. CC1 was derived from a portion of the stalk domain of kinesin, adopting a tetrameric coiled-coil structure. In contrast, CC2 was designed by the Kopeček group as a tetrameric coiled-coil, but comprised a shorter sequence than CC1. Importantly, it was observed that CC2 exhibited greater thermostability than CC1. These investigators hypothesized that differences in amino acid sequence between the two coiled-coil motifs (CC1 and CC2) would produce measurable effects in subsequent downstream biophysical properties. The difference in sequence between CC1 and CC2 influences the thermodynamic stability of the resulting coiled-coils, which, in turn, would affect the properties of the resultant hydrogel. For example, the T_m of a coiled-coil could be raised or lowered through sequence control, which enabled the formation of hydrogels of varying thermostability. When subjected to a temperature increase from 25° to 70 °C, the CC1-based hydrogel volume decreased tenfold (Wang et al. 1999). The mid-point temperature of the gel-sol transition coincided with the T_m of the constituent coiled-coil that was determined from its CD melting transition. Conversely, CC2 did not exhibit a hydrogel volume transition within the same temperature range (Wang et al.

1999). Kopeček asserted that this data indicated that "properties of a well-defined coiled-coil protein motif can be imposed onto a hybrid hydrogel" (Kopeček 2007).

In the pursuit of enhanced stability and a dynamic mechanical response, Yao et al. created a hybrid hydrogel that incorporated reversible physical cross-links and photo-polymerizable groups (Yao et al. 2014). A reactive cysteine residue was introduced into the sequence of a coiled-coil derived from the five-helix bundle of cartilage oligomeric matrix protein (COMP). The COMP peptide was conjugated to a macromer of polyethylene glycol diacrylate (PEGDA) through Michael addition to one of the terminal acrylate groups. The resultant pentameric assembly contained PEG chains that were terminally functionalized with the remaining acrylate group. Photo-induced polymerization of the terminal acrylates afforded a network in which the pentameric coiled-coil acted as a physical cross-link between polymer chains. This process has the advantage of retaining the dynamic physical cross-links associated with the coiled-coil motif, while permitting rapid *in situ* fabrication of the hydrogel using photo-induced polymerization (Yao et al. 2014). Fibroblast cells could be encapsulated within a 3D hydrogel matrix, which was generated *in situ* via photo-polymerization of an RDG-modified peptide-macromer. The reversibility of the coiled-coil self-association enabled self-healing behaviour within the hydrogel, which promoted cell spreading and migration in 3D culture (Yao et al. 2014). These hydrogel systems demonstrate a key advantage in the use of coiled-coils as structural units within self-assembling biomaterials, namely the potential for customization of the selectivity and stability of the interaction.

In the cases described above, the coiled-coil was employed primarily as a biorecognition motif in order to create the virtual cross-links that defined the hydrogel. The hydrophilic central block provided the flexibility to accommodate the swelling solvent. However, Woolfson and co-workers demonstrated hydrogels could be created solely from α-helical peptide sequences, which they termed hydrogelating self-assembling fibres (hSAFs) (Banwell et al. 2009). The hSAF system is based on previous work performed by the Woolfson group, wherein they engineered self-assembling fibres (SAFs) via two α-helices co-assembling such that the resultant structure was a dimer of the two helices. Importantly, this dimer featured sticky ends, and resultant end-to-end assembly produced coiled-coil fibrils, which in turn aggregated to form large fibres. It was observed that these fibres demonstrated a high level of order, leading to their precipitation from solution (Banwell et al. 2009). Using the SAF as a starting scaffold, the researchers inserted alanine residues (to encourage hydrophobic interactions between fibrils) and glutamine residues (to promote hydrogen bonding). These mutations were performed only at the *b, c,* and *f* positions of the heptad repeat, as these lie on the exposed surface of the helix, while other residue positions are more critical in driving dimerization of the two constituent helices to form a coiled-coil. Differing levels of alanine and glutamine were introduced at the aforementioned positions to form hydrogels driven largely by hydrophobic interactions between fibrils or driven by hydrogen bonding. It was observed that the hydrophobically-driven hydrogels (those featuring proportionally large amounts of alanine at the variable positions) showed an increase in gel strength

when heated. Conversely, the hydrogel variants driven primarily by hydrogen bonding (those featuring proportionally large levels of glutamine at the variable positions) melted and lost their gel strength upon heating. A distinct advantage of this system is that each constituent helix has its own sequence identity, and hydrogelation only occurs upon mixing. Lastly, Woolfson and co-workers showed that these hSAF systems are compatible with, and support both cell growth and cell differentiation. A recurring theme regarding coiled-coil biomaterials is the flexibility of this protein fold as a structural platform. Mehrban et al. demonstrated that these fibrillar coiled-coils could be decorated with the RGDS cell-adhesive sequence. The resultant hydrogels could promote the growth, directional migration, and differentiation of murine embryonic neural stem cells in culture (Banwell et al. 2009; Mehrban et al. 2015).

17.4 Immunogenic Coiled-Coil Assemblies

An important factor to consider in the development of biomaterials for *in vivo* applications is immunogenic response. For tissue engineering applications, minimizing the immune response to the coiled-coil construct facilitates integration into the body with minimal risk of rejection. In contrast, immunotherapies rely on a strong immune response to the coiled-coil material (Boato et al. 2007). It follows that it is of great interest to not only measure the immunogenic potential of coiled-coil assemblies under physiologically relevant conditions for *in vivo* applications, but also to design systems in which immunological response can be controlled.

For a biomaterial to provoke an immune response, it must interact with the immune system either by being recognized as a foreign body or by breaking down the native B-cell tolerance (if it is sufficiently homologous to a protein in the host organism). In both cases, aggregated (Maas et al. 2007) or oligomerized (Rosenberg 2008) proteins have been shown to be more immunogenic than their monomeric counterparts. Biomaterials that do not provoke an immune response are invisible to the host cell and function freely within the cell. Because coiled-coil systems can access a variety of oligomeric states, they are ideal for tuning an immune response.

Recently the Collier group explored the effect of PEGylation on the immunogenicity of coiled-coil assemblies. PEGylation generally reduces immune response, while aggregation increases the immune response (Rudra et al. 2010a). The interesting question then, is how strong will the immune response be when a PEGylated peptide aggregates? To address this question, they investigated the murine immune response to three peptides: a domain from mouse fibrinogen (which disfavours coiled-coil formation due to poor charge pairings and buried polar residues), a mutated version of the same with a coiled-coil optimized sequence (γKEI), and a triblock peptide-polymer conjugate (γKEI-PEG-γKEI) (Rudra et al. 2010b). They found that the native peptide and its coiled-coil derivative elicited no immune response, while the triblock peptide copolymer caused a mild immune response.

They attributed this change in immune response to the higher degree of oligomer-
ization observed in the triblock copolymer, as determined from analysis using ana-
lytical ultracentrifugation (AUC). These results suggested that the immunogenicity
of the peptide was not based on the folding, as the native random coil peptide and
the coiled-coil forming mutant elicited similarly low immune responses.

In a particularly interesting study, the Burkhard group developed a coiled-coil
nanoparticle with antiviral properties against Severe Acute Respiratory Syndrome
(SARS) (Pimentel et al. 2009). SARS is a pneumonia-like disease that caused wide-
spread concern in 2002, when it rapidly spread among 29 different countries and
resulted in the deaths of 800 people (Marra et al. 2003; Nicholls et al. 2003). The
epicentre of this outbreak was in southern China, where several domesticated ani-
mal species were found to be carrying the disease in live animal markets (Guan
et al. 2003). The search for a SARS vaccine resulted in the identification of the
SARS corona virus S protein. Attempts to use the full-length S protein to vaccinate
different model animals resulted in increased virulence of hepatitis in ferrets (Czub
et al. 2005) and infection with SARS in cats (Corapi et al. 1992). Thus, the design
of an S protein B cell epitope became an important research target. The key to
designing an effective epitope is mimicking the native protein structure; because of
this consideration, the coiled-coil forming region of the S protein was used as the
vaccine nanoparticle scaffold (Raman et al. 2006). They were able to develop a
monomer that would self-assemble into a viral capsid-mimicking icosahedral
nanoparticle (Fig. 17.5). The coiled-coil assembly had the advantages of high stabil-
ity, ease of expression, and lack of infectivity seen in the full S protein. In addition,
the coiled-coil assembly was found to be highly immunogenic due to its size, mim-
icry of the native epitope, and repetitive epitope display, making it a reasonable first
step to developing a SARS vaccine.

17.5   Coiled-Coil Based Nanotubes, Fibrils, and Fibres

One-dimensional assemblies derived from proteins and peptides have been the
focus of significant scientific research over the last two decades (Scanlon and Aggeli
2008). A significant motivation for this research stems from the presence of extended
helical assemblies in native biological systems, in which they display a variety of
functional roles that would be desirable to replicate within synthetic biomaterials.
The knobs-into-holes packing typically associated with coiled-coils has been
observed to play a significant role in stabilizing the structures of helical assemblies
such as intermediate filaments (Chernyatina et al. 2015), type IV pili (Craig et al.
2006), bacterial flagella (Yonekura et al. 2003), the type III secretion system needle
complex (Loquet et al. 2012), and filamentous phage capsids (Morag et al. 2015).
Significant research effort has been directed toward the *de novo* design of helical
assemblies based on coiled-coil structural motifs, which could be employed for the
creation of synthetic scaffolds for biomedical applications (see above, for example).
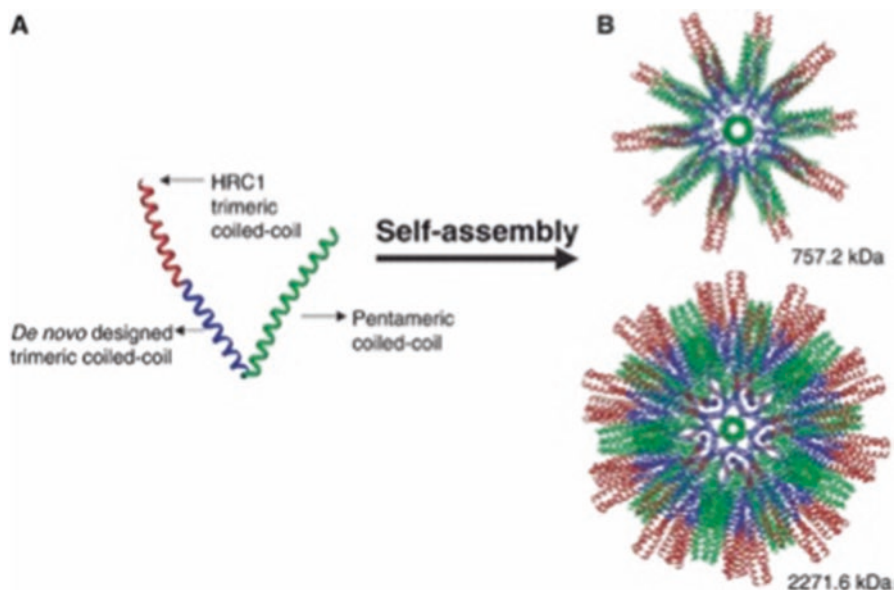
Fig. 17.5 Design of the icosahedral coiled coil epitope as a potential SARS vaccine. (**a**) the HRC1 domain and self assembly partners. (**b**) A 60 unit icosahedral assembly of the peptide nanoparticle (*top*) and a larger 180 unit icosahedral form (*bottom*) (Reprinted from Pimentel et al. (2009) Copyright 2009 *Chem Biol Drug Des*)

The primary approach of this research focused on design of sticky-ended coiled-coils that could oligomerize to form infinite one-dimensional polymers through non-covalent interactions. The staggered alignment of helices was promoted through introduction of electrostatic interaction between helical protomers that enforced an out-of-register alignment across the helical interface. However, most designs focused on coiled-coils of lower oligomerization state (usually dimers and trimers), primarily due to the difficulty in controlling alignment between adjacent helices for coiled-coils in larger helical bundles. In addition, at that juncture, limited structural information was available on coiled-coils of higher oligomerization state. One attractive feature of the latter systems is that coiled-coils of oligomerization state ≥ 5 helices display an interior channel in which the diameter depends on the bundle size. Self-assembly of coiled-coils of higher oligomerization state affords the potential to create nanotube structures in which the channel residues can be tailored for encapsulation of small molecules and the outer surface can be functionalized to enhance applications such as selective targeting for controlled release.

Numerous strategies have been described to drive the formation of tubular assemblies from peptides and proteins, though few of these approaches are applicable to coiled-coil systems (Burgess et al. 2015). The most promising approach to nanotube fabrication involves the end-to-end stacking of coiled-coils of defined oligomerization state ($n \geq 5$) into extended structures through selective introduction of attractive non-covalent interactions at the interfaces between subunits. As

coiled-coils are perhaps the best understood protein fold, with a wealth of available structural information, they represent a promising test bed to develop methods to reliably generate structurally-defined supramolecular assemblies.

Xu et al. (2013) described an approach to promote the end-to-end self-assembly of a seven-helix bundle into a helical nanotube. These researchers employed a structurally characterized 7-helix bundle, GCN4-pAA, as the starting point for the design. In this case, the 7-helix bundle has helical symmetry, in that successive peptides in the assembly are axially translated by an increment of one residue (1.43 Å rise per residue for the superhelix) between successive helices in the bundle. This translation results in a registry shift corresponding to a heptad repeat between the first and last helix in the bundle. The corresponding edges at the *N*- and *C*-termini are solvent-exposed and, upon interaction, could bury sufficient hydrophobic surface area to promote stacking of the 7-helix bundles. Xu et al. (2013) used the analogy of a helical lock washer to describe the structure of the coiled-coil subunits. The sequence of GCN4-pAA was redesigned to afford a fibrillogenic peptide sequence, 7HSAP1. This sequence preserved the residues that were critical for formation of the 7-helix bundle, but introduced features that would promote stacking of the subunits (Xu et al. 2013). Hydrophobically-driven association of complementary helix faces yielded a non-covalent assembly of helices with an inner lumen. In addition, the *N*-terminus and *C*-terminus of 7HSAP1 were populated with positively charged residues and negatively charged residues, respectively. This design promoted charge complementary association of helical heptamers through head-to-tail stacking of the heptameric subunits. The resultant nanotubes defined an internal cavity of approximately 7 Å in diameter, in analogy with the crystal structure of GCN4-pAA, which was capable of binding shape-appropriate small molecules such as the solvatochromic fluorophore PRODAN (Xu et al. 2013).

Hume et al. constructed a fibrillar assembly using the pentameric coiled-coil region of cartilage oligomeric matrix protein (COMPcc) as the basis for design. A variant of COMPcc was constructed in which residues that were non-essential for pentameric assembly were deleted (Hume et al. 2014). An *N*-terminal pocket, inherent to the wild type matrix protein and consisting of three leucines and a valine, was shifted with respect to the overall sequence, but preserved in identity and conserved with respect to its occurrence within the heptad repeat sequence. The presence of the *N*-terminal pocket was necessary to retain the preference for the pentameric assembly (over alternative oligomeric states or lack of assembly). Moreover, in the re-designed peptide, Q, charge density was introduced at positions on the outer (lateral) surface of the assembly in order to promote complementary interactions between oppositely charged regions on adjacent pentameric assemblies. Fibre formation would be promoted through head-to-tail stacking of helical bundles, which would be reinforced through lateral association between fibrils mediated through electrostatic complementarity. Peptide Q was found to form nano-fibres having diameters as large as half a micron (Hume et al. 2014). The pentameric assembly also defined a central pore of sufficient size to accommodate small molecules, such as the therapeutic agent curcumin. Curcumin was demonstrated not only to bind to

the fibrillar assemblies, but also to promote further lateral association through a mechanism that has yet to be fully elucidated. Using this approach, Hume et al. succeeded in engineering the first documented experimentally-assembled protein-based microfibre, observing diameters of up to sixteen microns via aggregation of nano-scale moieties, while preserving a central cavity for encapsulation (Hume et al. 2014).

Burgess et al. recently described a more general approach to the design of fibrils, including nanotubes, derived from coiled-coil sequences having different oligomerization states (Burgess et al. 2015). Helical bundles of defined oligomerization state ranging from three to seven were designed, in which residues at the *N*- and *C*-termini were chosen to be electrostatically complementary in charge (as described above). This design element was introduced to promote end-to-end association of coiled-coil bundles, affording one-dimensional assemblies from blunt-ended stacking of helical bundles. The resultant sequences (termed CC-Tri through CC-Hept), self-assembled into fibrils based on subunits having the corresponding oligomerization state, albeit with differing degrees of lateral association and order (Fig. 17.6). One system, CC-Hex-T, based on a hexameric helical bundle, exhibited a high degree of para-crystalline order after thermal annealing in PBS buffer. Burgess et al. further demonstrated that the hydrophobic dye 1,6-diphenylhexatriene was able to penetrate and bind within the inner lumen of tubular assemblies derived from CC-Hex-T, but not to the fibres derived from the tetrameric bundle CC-Tet2-F (Burgess et al. 2015). Subsequent experiments established that dye binding only occurred for helical bundles having an oligomerization state from 5–7, as these larger coiled-coil subunits define an inner lumen of sufficient size to accommodate the guest molecule. The encapsulation of 1,6-diphenylhexatriene validates the potential of these tubular materials for applications involving drug or small molecule binding and release.

While larger coiled-coil bundles (>7 subunits) have been observed in native protein assemblies, the *de novo* design of larger diameter coiled-coils remains a challenge, which has thus far limited the fabrication of nanotubes with larger lumens. Walshaw and Woolfson (Walshaw and Woolfson 2001) described general principles for design of these types of assemblies based on an analysis of the 12-helix bundle observed in the structure of the tolC homotrimer (Koronakis et al. 2000). However, such large diameter tubular structures based on coiled-coils have yet to be realized. Egelman et al. (2015) investigated the self-assembly of a pair of *de novo* designed coiled-coils based on the structural principles proposed by Walshaw and Woolfson. The sequences of these two peptides were based on a canonical heptad repeat motif in which hydrophobic residues populate the *a/d* and *c/f* positions, producing hydrophobic seams along the helix offset by two residue spacing (Koronakis et al. 2000). In contrast to the α-helical cylinder of the tolC sequence (Calladine et al. 2001) hydrophobic residues at the outer *a* and *f* positions were substituted with alanines, which feature smaller sidechains than the residues of the native tolC. This substitution served to decrease steric interference between the flanking residues of the two hydrophobic pockets to encourage larger tube diameter (Egelman et al. 2015). Large sidechains at the outer positions would create steric interactions between neigh-
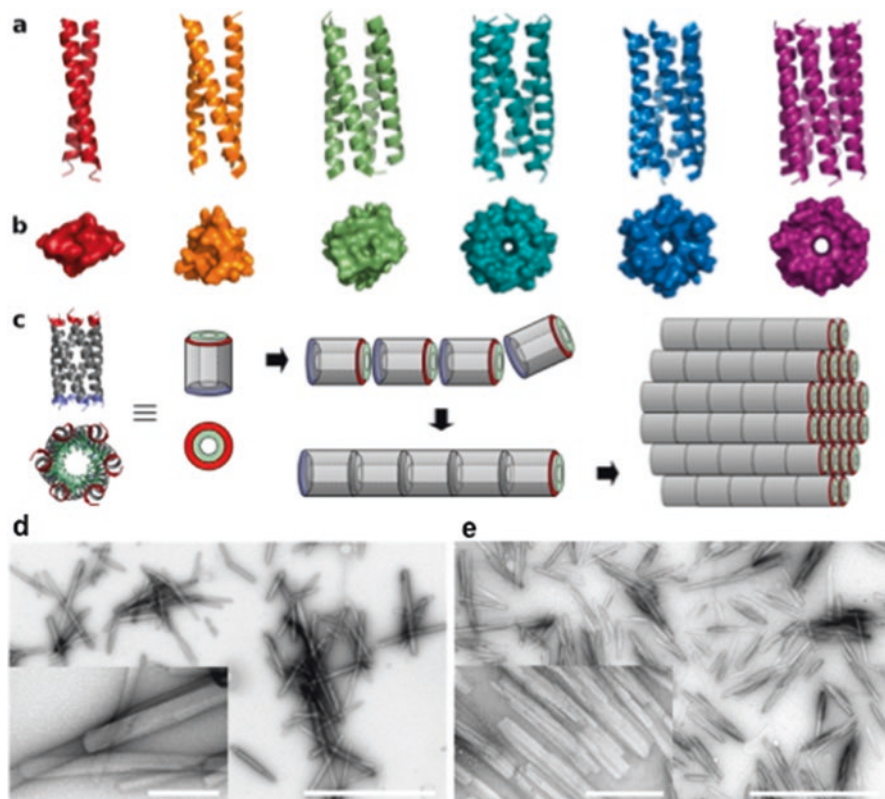
Fig. 17.6 [*top*] X-ray crystal-derived graphical representations of CC-Di through CC-Hex. [*bottom*] Transmission electron micrographs of tube assemblies (**a**) Graphical representation of coiled-coil oligomers ranging from two to seven, indicating the overall structure of the coiled-coil subunit as seen side-on. (**b**) X-ray crystal-derived graphical representations of coiled-coil oligomers ranging from two to seven indicating the overall structure of the coiled-coil subunit as seen when viewed along the fibre axis. In the pentamer, hexamer, and heptamer assemblies, a putative inner lumen is observed. (**c**) Putative method of assembly of coiled-coil subunits, wherein end-to-end association drives fibre formation. (**d**) Transmission electron micrographs of negative stained CC-Tet. (**e**) Transmission electron micrographs of CC-Pent (Reprinted with permission from Burgess et al. (2015) Copyright 2015 *J Am Chem Soc*)

bouring residues, and in the relaxation and mitigation of this steric stress, force the tube to constrict towards a smaller diameter. This was negated in the synthetic peptide via mutations to alanine at the aforementioned heptad positions *a* and *f* to encourage large diameter growth. Moreover, complementary pairs of charged residues populate the *b* and *e* positions in the synthetic peptide to promote antitypic helix-helix association (Egelman et al. 2015). The sole difference between the two sequences was the substitution of four arginine residues in the first peptide (Form I) with four lysine residues in the second peptide (Form II). Both peptides formed

extended tubular assemblies in aqueous buffer, but differed significantly in diameter (6 nm for the Form I assemblies and 12 nm for the Form II assemblies).

Electron cryomicroscopy with direct electron detection was employed for the structural analysis of the Form I and Form II assemblies. Egelman et al. (2015) funnelled data through the iterative helical real space reconstruction algorithm (Egelman 2010) to produce near atomic resolution structural data for Forms I and II. Form I features a unilaminar helical motif, wherein helices are packed nearly perpendicular to the nanotube axis in which the cross-section displays near four-fold symmetry in the left-handed one-start helix. The Form II assemblies feature a bilaminar helical motif, wherein three bilayer stacks of helices are packed together and define a nearly circular cross-section in which the helices are slightly inclined with respect to the cross-sectional plane. In either structure, the nanotube lumen and outer surface are lined with functional groups from polar amino acid residues. One might envision that the inner and outer surfaces of the tubes could be selectively functionalized to promote applications in controlled encapsulation and release.

Notably, the structures of the Form I and II assemblies diverged significantly from the structural model proposed by Walshaw and Woolfson, as well as that of the tolC α-cylinder (Koronakis et al. 2000; Calladine et al. 2001; Egelman et al. 2015). More remarkably, the structurally-conservative substitution of arginine for lysine resulted in significant differences in structure between the two helical assemblies. These differences highlighted a potential problem in the design of such assemblies, namely, that quaternary structure may not be robust in sequence space. Indeed, Egelman et al. demonstrated that a single mutation (R13K) was sufficient for conversion of the Form I assembly to the Form II assembly (Fig. 17.7). Conversely, a coupled pair of mutations (K13R, K17R) provides for conversion of Form II assembly to the Form I assembly (Egelman et al. 2015). These structural differences are manifested as a consequence of different local interactions between subunits that are propagated hierarchically to form the two distinct assemblies. The Form I assembly derives its stability from the interaction of the sidechains of Arg13 and Arg17 with the C-terminal residues of a helix in an adjacent stack. These interactions form the corners of the fourfold symmetric cross-section of the tubular assembly. In contrast, the helical stacks of Form II assembly, lacking these structural critical arginine residues, are held together weakly through interactions between the termini of helices in adjacent stacks. The structural changes that occur as a consequence of mutagenesis can be understood in terms of abrogation or introduction of these arginine-derived interactions (Egelman et al. 2015). The difference between the Form I and Form II assemblies dramatically illustrates that variations in primary sequence can result in large changes in quaternary structures. Atomic resolution structural information enables the identification of these critical interactions that underlie these differences, but the development of coiled-coils with high oligomerization state as biomaterials will require better methods to reliably design higher order structure.

In addition to tubular assemblies, coiled-coils have been employed in the fabrication of capsular and film-like biomaterials, which have advantages as carriers for encapsulation and controlled release of drugs and other small molecules. Utilizing
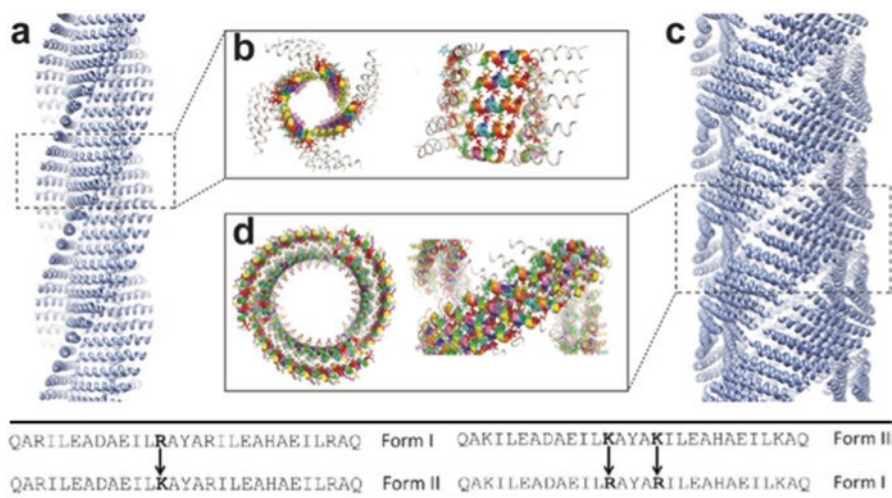
Fig. 17.7 [*top*] Atomic models for knobs-into-holes packing SOCKET analysis for Form I and Form II. [*bottom*] Sequences of Form I and II, illustrating the mutations required for inter-conversion between forms. (**a**) Graphical representation of atomic model for Form I via SOCKET analysis. (**b**) Graphical representation of knobs-into-holes packing of the C-terminus of Form I as seen [*left*] from along the fibre axis and [*right*] as seen from a side-on view. (**c**) Graphical representation of atomic model for Form II via SOCKET analysis. (**d**) Graphical representation of knobs-into-holes packing in Form II as seen [*left*] from along the fibre axis and [*right*] as seen from a side-on view. KIH packing is evidenced only within the same wall, and KIH packing does not occur between the two lumen (*inner* and *outer*) of the tube [(Egelman et al. 2015)] (Reprinted with permission from Egelman et al. (2015) Copyright 2015 *Structure*)

a layer-by-layer (LbL) approach, Gormley et al. have been able to construct cap-sules and films by subsequent addition of layers in a sequential manner (Gormley et al. 2015). Traditionally, deposition of sequential layers was achieved via self-assembly that was mediated through electrostatic complementarity between layers (Hammond 1999; Ai et al. 2003). Since this discovery, alternative means for LbL deposition have been accomplished in an effort to imbue enhanced levels of custom-ization and specificity with respect to biophysical properties of the assembly. The LbL toolbox contains other methodologies such as hydrogen bonding, covalent bonding, DNA-based oligomerization, and streptavidin-biotin complementarity (Mart et al. 2006). Stevens and co-workers have added a new methodology to the toolkit by using the specificity inherent in the self-assembly of coiled-coil moieties (Gormley et al. 2015). Using helix-loop-helix motifs (JR2EC and JR2KC) that oligomerize as a coiled-coil tetramer, these researchers created a polymer-peptide hybrid by conjugation of cysteine to *N*-(2-hydroxypropyl)methacrylamide-*co*-*N*-(3-aminopropyl)-methacrylamide [HPMA-*co*-APMA] polymer through a 4-(*N*-maleimidomethyl) cyclohexanecarboxylic acid *N*-hydroxysuccinimide ester [SMCC] linker (Gormley et al. 2015). The linkage moiety was attached to C22 at the loop region of the helix-loop-helix. Peptides JR2EC and JR2KC differ in sequence by a total of eight glutamates/lysines, which imbued a difference in net
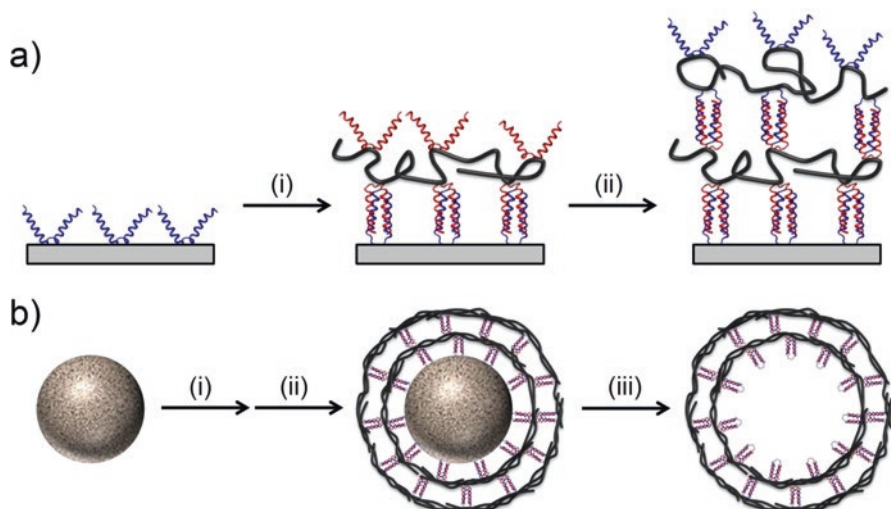
Fig. 17.8 Graphical representations of (**a**) planar LbL assembly and (**b**) spherical, colloidal LbL assembly. Tetramer coiled-coil formation drives the sequential peptide-polymer assembly of each layer (Reprinted with permission from Gormley et al. (2015) Copyright 2015 *Chem Mater*)

charge to each variant. This duality in charge promotes selective heteromerization of the two helix-loop-helix moieties, JR2EC and JR2KC, to form a tetrameric coiled-coil. As the peptide helices are conjugated to a polymer scaffold, coiled-coil formation results in sandwich-like constructions of iterative layers by sequential addition of polymer-peptide units with opposite charges. Anchoring techniques could be implemented to grow these LbL assemblies on both planar and spherical, colloidal substrates, giving a degree of geometric variation to what can be accomplished via this approach (Fig. 17.8). Lastly, as with many coiled-coil-based systems, a pH-induced conformational switch can be actuated. At pH 7, the dimer subunits stack against one another in the film. At pH 4, positive charges dominate the surface and the dimers repel one another, causing dissociation of the stacked units (Gormley et al. 2015).

17.6 Coiled-Coil Composites

This review has largely focused largely on developments related to the potential of coiled-coils for applications in biomedicine, and, specifically, their interactions with other biological materials *in vitro* or *in vivo*. Recently, however, significant research has been focused on integrating inorganic materials with coiled-coil biomaterials. In particular, the use of coiled-coils has been proposed for applications involving the templating of inorganic materials. For example, numerous studies have used genetically engineered viral capsids for directing the deposition of inorganic materials,
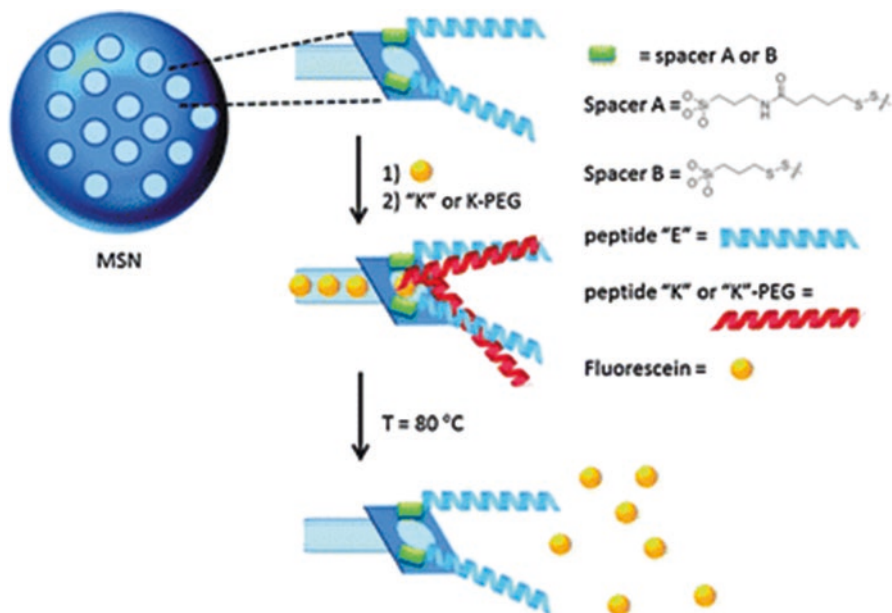
Fig. 17.9 A schematic representation of the mesoporous silica nanoparticle (MSN) system. The MSN construct is depicted in *blue*, with the small pores indicated by *light blue*. Spacers (*green squares*) of varying length are conjugated to the "E" peptide (*blue ribbon*), allowing it the flexibility needed to cover the pores. Addition of peptide "K" (*red ribbon*) leads to the sealing of the pore upon heterodimerization. Temperature increase leads to the unwinding of the coiled-coil, resulting in cargo (*yellow spheres*) delivery (Reprinted from Martelli et al. 2013)

leading to the formation of highly ordered inorganic structures (Lee et al. 2002; Huang et al. 2005). These strategies have produced very efficient electron-capture devices (Dang et al. 2011), as well as highly functioning lithium batteries (Lee et al. 2009). While using a phage capsid is an elegant solution to the selection problem, a coiled-coil scaffold would be preferable due to its ease of production, robust structure, and well-known design rules. Several examples have been recently published in which coiled-coils, or assemblies derived from their self-association, have been employed to direct nanoparticle synthesis and assembly (Ryadnov 2006; Slocik et al. 2007; Wagner et al. 2009; Hume et al. 2015; Abram et al. 2016). These materials might ultimately find use as imaging agents or therapeutics in medical applications, especially as they could leverage the favourable electronic, optical, or magnetic properties of the inorganic nanoparticle and the structural specificity of the coiled-coil interaction (Seo et al. 2015).

This objective was realized in a particularly innovative biomaterial design based on a coiled-coil composite. Martelli et al. developed a hybrid system that consisted of a coiled-coil heterodimer, fluorescein cargo, and mesoporous silica nanoparticles (Fig. 17.9) (Martelli et al. 2013). Briefly, mesoporous silica nanoparticles encapsulating large quantities of fluorescein were fabricated and conjugated to an acidic

coiled-coil peptide E. These complexes were then incubated overnight with the complementary basic coiled-coil peptide K, which was free in solution. The E and K peptides underwent selective heterodimerization, which resulted in *de facto* closure of the silica pores. The coiled-coil dimer could essentially perform as a valve, which could be actuated through denaturation of the dimer. For example, thermolysis of these complexes induced denaturation of the coiled-coil dimer. Dissociation of the K peptide actuated the valves and resulted in release of the cargo. Fluorescein was selected as a model cargo due to its fluorescent properties and ease of imaging. These materials have considerable potential for *in vivo* applications, especially if this approach proves to be more general. For example, ferromagnetic iron oxide nanoparticles would enable targeting to a specific location, e.g., a tumour, using an externally applied magnetic field. Once in the correct location, the cargo, e.g., a cytotoxic drug molecule could be released via localized heating in which the temperature leads to thermal denaturation of the heterodimeric coiled-coil.

17.7 Conclusion

Recent technological advances, such as structurally predictive computational algorithms and higher resolution characterization techniques, are expanding our ability to design and structurally characterize novel coiled-coil assemblies. The availability of this information has facilitated the development of novel biomaterials scaffolds based on coiled-coils. The natural abundance and structural diversity of coiled-coils, as well as their well-understood design rules, make them an obvious choice for biomaterials scaffolding. In addition, the versatility of the coiled-coil motif allows for the incorporation of artificial amino acids, which enhances the ability to do site-specific chemistry and generate more complex materials. As a result, coiled-coils have been used to develop a variety of structures, including hydrogels, nanotubes, nanosheets, and even composite materials with synthetic polymers or inorganic nanoparticles. Given the ever-increasing need for tailored materials, coiled-coils are emerging as a highly modular scaffold capable of a wide variety of functions.

References

Abram SB, Aupič J, Dražićd G, Gradišar H, Jerala R (2016) Coiled-coil forming peptides for the induction of silver nanoparticles. Biochem Biophys Res Commun 472(3):566–571
Ai H, Jones SA, Lvov YM (2003) Biomedical applications of electrostatic layer-by-layer nanoassembly of polymers, enzymes, and nanoparticles. Cell Biochem Biophys 39:23–43
Anzini P, Xu C, Hughes S, Magnotti E, Jiang T, Hemmingsen L, Demeler B, Conticello VP (2013) Controlling self-assembly of a peptide-based material via metal-ion induced registry shift. J Am Chem Soc 135(28):10278–10281

Apostolovic B, Deacon SPE, Duncan R, Klok H-A (2010) Hybrid polymer therapeutics incorporating bioresponsive, coiled coil peptide linkers. Biomacromolecules 11:1187–1195

Apostolovic B, Deacon SPE, Duncan R, Klok H-A (2011) Cell uptake and trafficking behavior of non-covalent, coiled-coil based polymer-drug conjugates. Macromol Rapid Commun 32:11–18

Assal Y, Mizuguchi Y, Mie M, Kobatake E (2015) Growth factor tethering to protein nanoparticles via coiled-coil formation for targeted drug delivery. Bioconjug Chem 26:1672–1677

Banwell EF, Abelardo ES, Adams DJ, Birchell MA, Corrigan A, Donald AM, Kirkland M, Serpell LC, Butler MF, Woolfson DN (2009) Rational design and application of responsive alpha-helical peptide hydrogels. Nat Mater 8:596–600

Boato F, Thomas RM, Ghasparian A, Freund-Renard A, Moehle K, Robinson JA (2007) Synthetic virus-like particles from self-assembling coiled-coil lipopeptides and their use in antigen display to the immune system. Angew Chem Int Ed Eng 46:9015–9018

Burgess NC, Sharp TH, Thomas F, Wood CW, Thomson AR, Zaccai NR, Brady RL, Serpell LC, Woolfson DN (2015) Modular design of self-assembling peptide-based nanotubes. J Am Chem Soc 137(33):10554–10562

Calladine CR, Sharff A, Luisi B (2001) How to untwist an α-helix: structural principles of an α-helical barrel1. J Mol Biol 305(3):603–618

Chernyatina AA, Guzenko D, Strelkov SV (2015) Intermediate filament structure: the bottom-up approach. Curr Opin Cell Biol 32:65–72

Chu T-W, Kopeček J (2015) Drug-free macromolecular therapeutics – a new paradigm in polymeric nanomedicines. Biomed Sci 3:908–922

Corapi WV, Olsen CW, Scott FW (1992) Monoclonal antibody analysis of neutralization and antibody-dependent enhancement of feline infectious peritonitis virus. J Virol 66:6695–6705

Craig L, Volkmann N, Arvai AS, Pique ME, Yeager M, Egelman EH, Tainer JA (2006) Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. Mol Cell 23(5):651–662

Czub M, Weingartl H, Czub S, He R, Cao J (2005) Evaluation of modified vaccinia virus ankara based recombinant SARS vaccine in ferrets. Vaccine 23:2273–2279

Dang X, Yi H, Ham MH, Qi J, Yun DS, Ladewski R, Strano MS, Hammond PT, Belcher AM (2011) Virus-templated self-assembled single-walled carbon nanotubes for highly efficient electron collection in photovoltaic devices. Nat Nanotechnol 6:377–384

Dublin SN, Conticello VP (2008) Design of a selective metal ion switch for self-assembly of peptide-based fibrils. J Am Chem Soc 130:49–51

Egelman EH (2010) Reconstruction of helical filaments and tubes. Methods Enzymol 482:167–183

Egelman EH, Xu C, DiMaio F, Magnotti E, Modlin C, Yu X, Wright E, Baker D, Conticello VP (2015) Structural plasticity of helical nanotubes based on coiled-coil assemblies. Structure 23(2):280–289

Elisseeff J (2008) Hydrogels: structure starts to gel. Nat Mater 7(4):271–273

Fletcher NL, Lockett CV, Dexter AF (2011) A pH-responsive coiled-coil peptide hydrogel. Soft Matter 7(21):10210–10218

Frampton JP, Hynd MR, Shuler ML, Shain W (2011) Fabrication and optimization of alginate hydrogel constructs for use in 3D neural cell cuture. Biomed Mater 6:18

Garbern JC, Minami E, Stayton PS, Murry CE (2011) Delivery of basic fibroblast growth factor with a pH-responsive, injectable hydrogel to improve angiogenesis in infarcted myocardium. Biomaterials 32:2407–2416

Gormley AJ, Chandrawati R, Christofferson AJ, Loynachan C, Jumeaux C, Artzy-Schnirman A, Aili D, Yarovsky I, Stevens MM (2015) Layer-by-layer self-assembly of polymer films and capsules through coiled-coil peptides. Chem Mater 27(16):5820–5824

Grigoryan G, Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. J Mol Biol 405(4):1079–1100

Gruber M, Lupas AN (2003) Historical review: another 50th anniversary – new periodicities in coiled coils. Trends Biochem Sci 28(12):679–685

Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302:276–278

Guziewicz N, Best A, Perez-Ramirez B, Kaplan DL (2011) Lyophilized silk fibroin hydrogels for the sustained local delivery of therapeutic monoclonal antibodies. Biomaterials 32:2642–2650

Hammond PT (1999) Recent explorations in electrostatic multilayer thin film assembly. Curr Opin Colloid Interface Sci 4:430–442

Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262(5138):1401–1407

Harbury PB, Kim PS, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. Nature 371:80–83

Huang Y, Chiang CY, Lee SK, Gao Y, Hu EL, De Yoreo J, Belcher AM (2005) Programmable assembly of nanoarchitectures using genetically engineered viruses. Nano Lett 5:1429–1434

Huang CC, Ravindran S, Yin Z, George A (2014) 3-D self-assembling leucine zipper hydrogel with tunable properties for tissue engineering. Biomaterials 35(20):5316–5326

Hume J, Sun J, Jacquet R, Renfrew PD, Martin JA, Bonneau R, Gilchrist ML, Montclare JK (2014) Engineered coiled-coil protein microfibers. Biomacromolecules 15(10):3503–3510

Hume J, Chen R, Jacquet R, Yang M, Montclare JK (2015) Tunable conformation-dependent engineered protein-gold nanoparticle nanocomposites. Biomacromolecules 16(6):1706–1713

Jung JP, Moyano JV, Collier JH (2011) Multifactorial optimization of endothelial cell growth using modular synthetic extracellular matrices. Integr Biol 3:185–196

Kopecek J (2007) Hydrogel biomaterials: a smart future? Biomaterials 28(34):5185–5192

Kopecek J, Yang J (2012) Smart self-assembled hybrid hydrogel biomaterials. Angew Chem Int Ed Eng 51(30):7396–7417

Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature 405(6789):914–919

Lee SW, Mao C, Flynn CE, Belcher AM (2002) Ordering of quantum dots using genetically engineered viruses. Science 296:892–895

Lee YJ, Yi H, Kim WJ, Kang K, Yun DS, Strano MS, Ceder G, Belcher AM (2009) Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. Science 324:1051–1055

Lin YH, Lin JH, Peng SF, Yeh CL, Chen WC, Chang TL, Liu MJ, Lai CH (2011) Multifunctional gentamicin supplementation of poly(gamma-glutamic acid)-based hydrogels for wound dressing application. J Appl Polym Sci 120:1057–1068

Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A (2012) Atomic model of the type III secretion system needle. Nature 486(7402):276–279

Lupas AN (1996) Coiled coils: new structures and new functions. Trends Biochem Sci 21(10):375–382

Lupas AN, Gruber M (2005) The structure of α-helical coiled coils. Adv Protein Chem 70:37–78

Maas C, Hermeling S, Bouma B, Jiskoot W, Gebbink MFBG (2007) A role for protein misfolding in immunogenicity of biopharmaceuticals. J Biol Chem 282:2229–2236

Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, Mcdonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andronov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray

M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Krajden M, Petric M, Skowronski DM, Upton C, Roper RL (2003) The genome sequence of the SARS-associated coronavirus. Science 300:1399–1404

Mart RJ, Osborne RD, Stevens MM, Ulijn RV (2006) Peptide-based stimuli-responsive biomaterials. Soft Matter 2(10):822

Martelli G, Zope HR, Capell MB, Kros A (2013) Coiled-coil peptide motifs as thermoresponsive valves for mesoporous silica nanoparticles. Chem Commun 49:9932–9934

Mehrban N, Zhu B, Tamagnini F, Young FI, Wasmuth A, Hudson KL, Thomson AR, Birchall MA, Randall AD, Song B, Woolfson DN (2015) Functionalized α-helical peptide hydrogels for neural tissue engineering. ACS Biomater Sci Eng 1(6):431–439

Morag O, Sgourakis NG, Baker D, Goldbourt A (2015) The NMR–rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. Proc Natl Acad Sci 112(4):971–976

Nicholls J, Dong XP, Jiang G, Peiris M (2003) SARS: clinical virology and pathogenesis. Respirology 8:6–8

Pechar M, Pola R, Laga R, Ulbrich K, Bednárová L, Maloň P, Sieglová I, Král V, Fábry M, Vaněk O (2011) Coiled coil peptides as universal linkers for the attachment of recombinant proteins to polymer therapeutics. Biomacromolecules 12:3645–3655

Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA (1998) Reversible hydrogels from self-assembling artificial proteins. Science 281(5375):389–392

Pimentel TA, Yan Z, Jeffers SA, Holmes KV, Hodges RS, Burkhard P (2009) Peptide nanoparticles as novel immunogens: design and analysis of a prototypic severe acute respiratory syndrome vaccine. Chem Biol Drug Des 73:53–61

Pola R, Laga R, Ulbrich K, Sieglová I, Král V, Fábry M, Kabešová M, Kovář M, Pechar M (2013) Polymer therapeutics with a coiled coil motif targeted against murine BCL1 leukemia. Biomacromolecules 14:881–889

Raman S, Machaidze G, Lustig A, Aebi U, Burkhard P (2006) Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. Nanomedicine 2:95–102

Rosenberg AS (2008) Effects of protein aggregates: an immunologic perspective. AAPS J 8:E572

Rudra JS, Tian PK, Hildeman DA, Jung JP, Collier JH (2010a) Immune responses to coiled coil supramolecular biomaterials. Biomaterials 31:8475–8483

Rudra JS, Tian YF, Jung JP, Collier JH (2010b) A self-assembling peptide acting as an immune adjuvant. Proc Natl Acad Sci U S A 107:622–627

Ryadnov MG (2006) A self-assembling peptide polynanoreactor. Angewandte Chemie Int Ed 46(6):969–972

Scanlon S, Aggeli A (2008) Self-assembling peptide nanotubes. Nano Today 3(3-4):22–30

Seo JW, Ang J, Mahakian LM, Tam S, Fite B, Ingham ES, Beyer J, Forsayeth J, Bankiewicz KS, Xu T, Ferrara KW (2015) Self-assembled 20-nm 64Cu-micelles enhance accumulation in rat glioblastoma. J Control Release 220:51–60

Shen W, Zhang K, Kornfield JA, Tirrell DA (2006) Tuning the erosion rate of artificial protein hydrogels through control of network topology. Nat Mater 5(2):153–158

Slaughter BV, Khurshid SS, Fisher OZ, Khademhosseini A, Peppas NA (2009) Hydrogels in regenerative medicine. Adv Mater 21(32-33):3307–3329

Slocik JM, Tam F, Halas NJ, Naik RR (2007) Peptide-assembled optically responsive nanoparticle complexes. Nano Lett 7(4):1054–1058

Testa OD, Moutevelis E, Woolfson DN (2009) CC+: a relational database of coiled-coil structures. Nucleic Acids Res 37(suppl 1):D315–D322

Thanasupawat T, Bergen H, Hombach-Klonisch S, Krcek J, Ghavami S, Del Bigio MR, Krawitz S, Stelmack G, Halayko A, McDougall M, Meier M, Stetefeld J, Klonisch T (2015) Platinum (IV) coiled coil nanotubes selectively kill human glioblastoma cells. Nanomedicine 11:913–925

Thompson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble α-helical barrels. Science 346(6208):485–488

Wagner SC, Roskamp M, Cölfen H, Böttcher C, Schlecht S, Koksch B (2009) Switchable electrostatic interactions between gold nanoparticles and coiled coil peptides direct colloid assembly. Org Biomol Chem 7:46–51

Walshaw J, Woolfson DN (2001) Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. Protein Sci A Public Protein Soc 10(3):668–673

Wang C, Stewart RJ, Kopecek J (1999) Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. Nat Lett 397:417–421

Wong VW, Rustad KC, Galvez MG, Neofytou E, Glotzbach JP, Januszyk M, Major MR, Sorkin M, Longaker MT, Rajadas J, Gurtner GC (2011) Engineered pullalan-collagen composite dermal hydrogels improve early cutaneous wound healing. Tissue Eng 17:631–644

Wood CW, Bruning M, Ibarra AA, Bartlett GJ, Thomson AR, Sessions RB, Brady RL, Woolfson DN (2014) CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. Bioinformatics 30(21):3029–3035

Wu K, Yang J, Liu J, Kopeček J (2012) Coiled-coil based drug-free macromolecular therapeutics: in vivo efficacy. J Control Release 157:126–131

Xu C, Kopeček J (2008) Genetically engineered block copolymers: influence of the length and structure of the coiled-coil blocks on hydrogel self-assembly. Pharm Res 25(3):674–682

Xu C, Liu R, Mehta AK, Guerrero-Ferreira RC, Wright ER, Dunin-Horkawicz S, Morris K, Serpell LC, Zuo X, Wall JS, Conticello VP (2013) Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. J Am Chem Soc 135(41):15565–15578

Yao M-H, Yang J, Du M-S, Song J-T, Yu Y, Chen W, Zhao Y-D, Liu Y-D (2014) Polypeptide-engineered physical hydrogels designed from the coiled-coil region of cartilage oligomeric matrix protein for three-dimensional cell culture. J Mat Chem B 2(20):3123

Yonekura K, Maki-Yonekura S, Namba K (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature 424(6949):643–650

Yu YB (2002) Coiled-coils: stability, specificity, and drug delivery potential. Adv Drug Deliv Rev 54(8):1113–1129

Zimenkov Y, Dublin SN, Ni R, Tu RS, Breedveld V, Apkarian RP, Conticello VP (2006) Rational design of a reversible pH-responsive switch for peptide self-assembly. J Am Chem Soc 128:6770–6771

Chapter 18
Bioengineered Collagens

Barbara Brodsky and John A.M. Ramshaw

Contents

Abstract There is a great deal of interest in obtaining recombinant collagen as an alternative source of material for biomedical applications and as an approach for obtaining basic structural and biological information. However, application of recombinant technology to collagen presents challenges, most notably the need for post-translational hydroxylation of prolines for triple-helix stability. Full length recombinant human collagens have been successfully expressed in cell lines, yeast, and several plant systems, while collagen fragments have been expressed in *E. coli*. In addition, bacterial collagen-like proteins can be expressed in high yields in *E. coli* and easily manipulated to incorporate biologically active sequences from human collagens. These expression systems allow manipulation of biologically active sequences within collagen, which has furthered our understanding of the relationships between collagen sequences, structure and function. Here, recombinant

B. Brodsky (✉)
Department of Biomedical Engineering, Tufts University, Medford, MA, USA
e-mail: barbara.brodsky@tufts.edu

J.A.M. Ramshaw
CSIRO Manufacturing, Bayview Avenue, Clayton, VIC 3169, Australia
e-mail: john.ramshaw@csiro.au

studies on collagen interactions with cell receptors, extracellular matrix proteins, and matrix metalloproteinases are reviewed, and discussed in terms of their potential biomaterial and biomedical applications.

Keywords Recombinant collagen • Collagen interactions • Collagen triple-helix • Proline hydroxylation • Bacterial collagens • Biomaterials

Abbreviations

CL	collagen domain of the bacterial protein Scl2
D	67 nm axial periodicity
DDR	discoidin domain receptor
ECM	extracellular matrix
EGF	epidermal growth factor
Fn	fibronectin
IL-2	interleukin-2
MMP	matrix metalloproteinase
P4H	prolyl 4-hydroxylase
Scl2	Streptococcus pyogenes collagen-like protein 2
Tm	melting temperature
V	trimerization domain of Scl2
vWF	von Willebrand factor

18.1   Introduction

Collagen constitutes the most abundant protein in the human body, and serves as a natural material for a wide variety of biomedical and bioengineering applications. Many structural and physical properties of collagen fibrils are desirable for biomaterials, including mechanical strength, hydration, integrity and resistance to most common enzymes. In addition, the physiological interactions of collagen with cells and other components in the extracellular matrix (ECM) present opportunities for the generation of biomaterials with the capacity for natural biological interfaces. Inexpensive and relatively pure collagen for a wide range of applications can be obtained from animal sources, such as bovine hide, and animal tissues will continue to be the major source of collagen for biomaterials in the foreseeable future. Over the past few decades, there has been increasing interest in applying molecular biology to obtain collagen products. The expression of recombinant collagen proteins leads to higher purity than available through protein extraction. A very important advantage of recombinant collagen is the ability to manipulate the amino acid sequence. Directed modification of animal collagen sequences allows for the design

of targeted collagen products for specific applications and provides an important avenue to investigate basic features of collagen structure and its interactions. This review will focus on the use of recombinant collagens as a tool to study the structure and function of collagens using biochemical and biophysical techniques, and will also explore the potential applications of these modified molecules in biomaterials and biomedical engineering.

18.2   Background: Collagens

Collagens are defined as ECM molecules which contain the triple-helix molecular structure (Brodsky and Ramshaw 1997). The triple-helix comprises three separate polypeptide chains, each in a left-handed polyproline II-like helix, that are wound together into a tight, right-handed superhelix (Ramachandran and Kartha 1955; Rich and Crick 1961; Bella et al. 1994). A consequence of the structure of this tightly wound conformation is that every third residue in a chain resides inside the triple-helix, where only the smallest amino acid Gly can be accommodated. This leads to a characteristic repeating amino acid sequence $(Gly\text{-}Xaa\text{-}Yaa)_n$ that is observed in the sequence of all triple-helix domains. In addition to close packing of the three chains, the triple-helix is stabilized by a high content of the imino acids proline (Pro) and hydroxyproline (Hyp), which promote the polyproline-II structure of the individual chains. During the biosynthesis of collagen chains, Pro residues incorporated into the Yaa position of the Gly-Xaa-Yaa repeating sequence are post-translationally modified by prolyl-4-hydroxylase (P4H) to produce Hyp (Fig. 18.1). The high Hyp content, often constituting more than 10 % of all residues, is a distinctive feature of animal collagens. The Hyp residues provide additional thermal stability to the triple-helix, beyond that of Pro residues (Rosenbloom et al. 1973), as a result of stereoelectronic effects and hydrogen bonding interaction with the
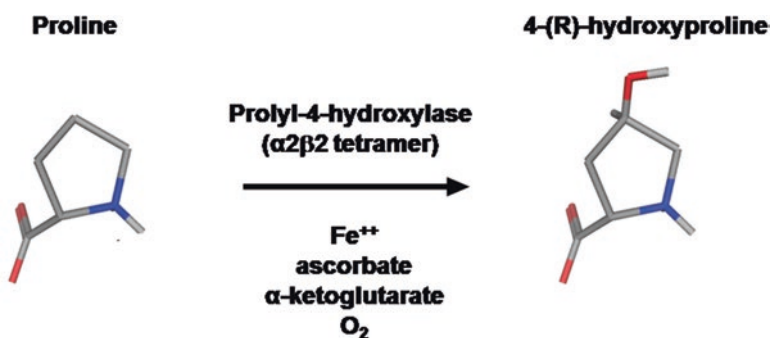
Fig. 18.1 Schematic diagram of enzymatic hydroxylation of Pro to 4-Hydroxyproline (Hyp) by the action of prolyl 4-hydroxylase (P4H) α2β2, and its requirements for ascorbate, iron, α-ketoglutarate and oxygen

hydration network (Shoulders and Raines 2009; Bella et al. 1995). In addition, Hyp has been implicated in the nucleation and efficiency of triple-helix folding (Ackerman et al. 1999), and may also be critical for collagen fibril formation (Kramer et al. 2000; Perret et al. 2001) and for collagen interactions with some receptors and ECM proteins (Smethurst et al. 2007; Perret et al. 2003).

In humans, the family of ECM proteins which contain triple-helix domains consists of 28 genetically-distinct collagen types (Ricard-Blum 2011). The most abundant are fibrillar collagens which form banded fibrils with an axial periodicity of 67 nm (D), such as type I collagen in skin, bone, cornea and tendon; type II collagen in cartilage; and type III collagen in skin and blood vessels. Chapter 14 (by Bella and Hulmes) in this volume, presents an in-depth review of these fibrillar collagens. The triple-helix domains of fibrillar collagens all contain a perfect (Gly-Xaa-Yaa)$_n$ repeat, and the essential nature of the Gly residues in these collagens is illustrated by the diseases observed when a mutation leads to the replacement of a single Gly residue within the triple-helix (Kuivaniemi et al. 1991; Byers and Cole 2002). In addition to the fibrillar collagens, there are ~20 types of non-fibrillar collagens. These include type IV collagen, which forms networks in basement membranes; transmembrane proteins such as types XV and XVII; and FACIT collagens, which bind to the surface of collagen fibrils (Ricard-Blum 2011). All non-fibrillar collagens contain one or more interruptions in the repeating tripeptide sequence pattern (Brodsky et al. 2008). For instance, type IV collagen chains have more than 20 sites where the Gly-Xaa-Yaa pattern is interrupted in their long triple-helix domain. Some triple-helix molecules are formed from three identical chains, such as the homotrimeric type II [α1(II)]$_3$ and type III [α1(III)]$_3$ collagens, while other collagens are heterotrimers. The most abundant collagen, type I, has two distinct chains which form the heterotrimer [α1(I)$_2$α2(I)]. Collagens are always synthesized with non-collagenous domains adjacent to the triple-helix domain(s), which may be important for trimerization, chain registration, assembly, or biological interactions (Boudko et al. 2012). Some of the non-collagenous domains are retained in the extracellular matrix after deposition, while others may be proteolytically cleaved and exhibit independent functions (Ortega and Werb 2002).

Although human collagens have been the most extensively characterized, collagens are found throughout the animal kingdom, including many insects (Yasothornsrikul et al. 1997; Francois et al. 1980; Sutherland et al. 2013) and some of the most primitive animal species, such as Cnidarians (hydra, jellyfish) (Fowler et al. 2000; Song et al. 2006) and certain Porifera (sponges) (Swatschek et al. 2002). In the last 10–15 years, a new group of triple-helical collagen-like proteins has been described from bacteria, based principally on database searches of bacterial genomes (Rasmussen et al. 2003). As described in a recent review, a number of these bacterial collagen-like proteins have been expressed in *E. coli* and shown to form stable triple-helices despite the lack of Hyp (Yu et al. 2014). One of the collagen-like proteins from *Streptococcus pyogenes*, Scl2, has been modified and studied as a means of probing triple-helix structure and collagen interactions and will be discussed at length in this review. Another group of collagen-like proteins has been reported in bacteriophage sequences embedded within various pathogenic strains of *E. coli*, and

these proteins, which have been shown to be triple-helical, may play a role in the attachment of phage particles to *E. coli* cells (Ghosh et al. 2012).

Most investigations of collagen structure and biological function have utilized extracted animal collagens. It is straightforward to obtain the major fibrillar collagens type I, II and III from the tissues of cows, pigs, and other animals in a high degree of purity, and these are now commercially available from a number of different sources (Abraham et al. 2008). The early structural studies on collagen, including fibre diffraction, electron microscopy, and hydrodynamic measurements, were carried out on extracted collagens. However, the long, rod-like extracted collagens did not prove amenable to more advanced molecular resolution techniques, such as X-ray crystallography and NMR. Peptides with Gly as every third residue and a high imino acid content are known to adopt a triple-helical conformation, and studies on collagen model peptides have played a major role in obtaining X-ray and NMR high resolution data on the triple-helical structure, as well as allowing correlations between sequence and stability (Bella et al. 1994; Buevich and Baum 2001; Persikov et al. 2000, 2002, 2005a, b; Brodsky and Persikov 2005). With the ability to express recombinant animal collagens, it has become possible recently to produce material with modified collagen sequences to augment the peptide studies. The ability to manipulate amino acid sequences in expressed collagens opens up avenues for studies of basic triple-helix properties, collagen functional properties, and the effects of mutations which cause disease. In addition, these designed recombinant collagens show promise for targeted biomedical applications.

18.3   Recombinant Collagens and Expression Systems

As molecular biology techniques made protein expression more routine, investigators began to take on the expression of recombinant collagens (Geddis and Prockop 1993; Bulleid et al. 2000). Recombinant collagen technology presented challenges not encountered in the production of most other proteins. The highly efficient *E. coli* and yeast expression systems do not have the collagen post-translational enzyme machinery found in animal cells. The use of these efficient systems presents the significant hurdle of hydroxylating proline residues to achieve molecular stability, since unhydroxylated collagen has a T_m value more than 10 °C lower than native hydroxylated collagen (Merle et al. 2002). In some cases, this challenge has been addressed by insertion of the genes for P4H, together with the desired collagen gene. Animal P4H is an $\alpha_2\beta_2$ tetramer composed of two distinct types of subunits, and the enzyme has specific requirements for oxygen, ascorbic acid, Fe^{2+}, and α-ketoglutarate (Fig. 18.1). The two human prolyl hydroxylase genes have been inserted in yeast systems, in bacterial systems, and in tobacco plants, but adjustments and modifications have been required to improve the level of post-translation modification, to try to reach physiological Hyp values (Merle et al. 2002; Olsen et al. 2003; Tang et al. 2016). In addition to the central hydroxylation issue, the

highly repetitive structure and the frequent Gly and Pro residues in collagen can cause expression problems.

A wide selection of different expression systems for recombinant animal collagens have been explored (Table 18.1) and have been the subject of several reviews (for example: Bulleid et al. 2000; Olsen et al. 2003; Báez et al. 2005; Ruggiero and Koch 2008; Werkmeister and Ramshaw 2012; Ramshaw et al. 2014). Each system

Table 18.1 Examples of expression systems for recombinant collagens

System	Host	Collagen type	
(a) Expression in systems with endogenous P4H activity			
Cell culture systems	HT1080 cells	Type I α1	Geddis and Prockop (1993)
	CHO cells	Type IV	Fukuda et al. (1997)
	COS cells	*Torpedo marmorata* AChE	Krejci et al. (1991)
	HEK293 cells	Type X	Frischholz et al. (1998)
	D. melanogaster S2 cells	Type XXI (NC1-Type 1) Fragment	Li et al. (2005)
	Baculovirus	Type I	Tomita et al. (1997)
		Type II	Nokelainen et al. (1998)
		Type III	Tomita et al. (1995)
		Type IX	Pihlajamaa et al. (1999)
		Type XIII	Snellman et al. (2000)
Transgenic animals	Mice	Type I α1	Toman et al. (1999)
	Silkworm (*B. mori*)	Type III fragment (20 %)	Tomita et al. (2003)
(b) Expression in systems with no P4H activity			
Transgenic plants	Tobacco leaves	Type I α1	Ruggiero et al. (2000)
	Maize seed	Type I α1	Zhang et al. (2009)
	Rice seed	Type II peptide (residues 250–270)	Hashizume et al. (2008)
Microorganisms	*E. coli*	Type II CB8, CB10	Tang et al. (1999)
(c) Expression in systems with added P4H activity			
Transgenic plants	Tobacco leaves	Type I α1	Merle et al. (2002)
		Type I	Stein et al. (2009)
	Maize seed	Type I α1	Xu et al. (2011)
Microorganisms	*S. cereviseae*	Type I	Toman et al. (2000)
	P. pastoris	Type I & Type III	Myllyharju et al. (2000)
	E. coli	Type III	Rutschmann et al. (2014)

Table 18.1 (continued)

System	Host	Collagen type	
(d) Expression in systems where no P4H activity is required for stability			
Microorganisms	*E. coli*	*S. pyogenes* Scl1	Xu et al. (2002)
		S. pyogenes Scl2	Xu et al. (2002)
		B. anthracis BclA	Boydston et al. (2005)
		L. pneumophilia Lcl	Vandersmissen et al. (2010)
		C. perfringens	Xu et al. (2010)
		S. usitatus	Xu et al. (2010)
		R. palustris	Xu et al. (2010)
		Methylobacterium sp.	Xu et al. (2010)
		Bacteriophage ECs2717 (from *E. coli* O157:H7)	Ghosh et al. (2012)
		*N. oligospilus (*willow sawfly) SfC A, SfC B, SfC C	Sutherland et al. (2013)

has advantages and disadvantages in producing collagen suitable for structure and function studies. For example, use of animal cell lines, such as HT1080, provides fully hydroxylated collagen in low yields, using endogenous P4H activity. The small amounts produced are useful for some studies, but may be insufficient for biophysical studies or large scale applications. On the other hand, the use of microorganisms such as *Pichia pastoris* or *Saccharomyces cereviseae* can give good yields, but the extent of genetic manipulation required to get effective strains that incorporate functional P4H is significant. Extensive work has been done using *P. pastoris* expression to obtain fully hydroxylated type I collagen which can form characteristic collagen-banded fibrils (Olsen et al. 2003). A modular strategy was devised to express human type III collagen in *S. cerevisiae*, and refinements were used to bring Hyp levels derived from inserted P4H genes closer to those found in the native collagen (Chan et al. 2010; Que et al. 2015). In *E coli*, human P4H subunits have been successfully incorporated (Kersteen et al. 2004), and genes were manipulated to include an ascorbate- like source needed for P4H activity (Pinkas et al. 2011). This has led to successful expression of small collagen-like peptides and the triple-helix containing protein adiponectin in *E. coli*, as well as a less defined human collagen-like fragment (Pinkas et al. 2011; Ding et al. 2012; Tang et al. 2016). A large fragment of human type III collagen (119 Gly-Xaa-Yaa triplets plus the N- and C-terminal telopeptides) was co-expressed with mimivirus prolyl and lysyl hydroxylase in *E. coli*, and shown to result in the production of the collagen fragment containing Hyp and hydroxylysine (Rutschmann et al. 2014). A number of transgenic plant systems have been used to express type I collagen (Table 18.1), including tobacco and maize, where again the P4H genes are also incorporated
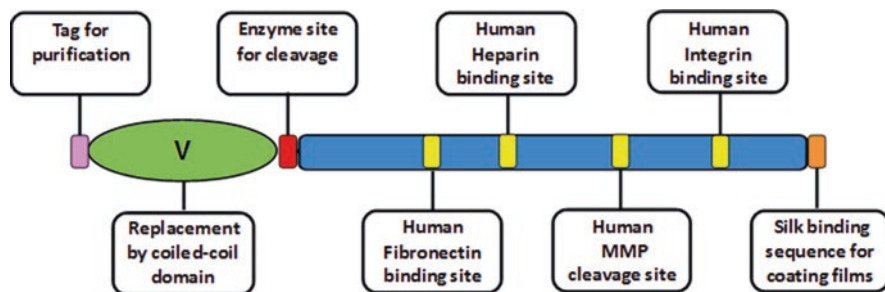
Fig. 18.2 Schematic representation of recombinant bacterial collagen construct, showing examples of sequence manipulations, including introduction of bioactive sites from human collagen, replacement of its trimerization domain by a coiled coil domain, and C-terminal addition of silk sequences

(Merle et al. 2002; Stein et al. 2009; Xu et al. 2011). However, the genetic manipulation and selection times needed for transgenic organisms could make their use challenging for studies where many different constructs are needed.

The emergence of bacterial collagen-like proteins has opened new windows of opportunity as the proteins are readily designed as synthetic genes and can be easily produced in good quantities as stable triple-helices in *E. coli* (Yu et al. 2014). The expressed bacterial collagen proteins have a stability close to that found in animal collagens despite the absence of Pro hydroxylation. The constructs are smaller and easier to manipulate than full length animal collagens, and they allow incorporation of larger test sequence elements than possible in peptide-based systems. Several recombinant bacterial collagen-like proteins have been well characterized, especially the Scl2 protein from *S. pyogenes*. The *S. pyogenes* collagen-like protein Scl2 construct contains an N-terminal trimerization domain (designated as V) and a (Gly-Xaa-Yaa)$_{79}$ collagenous domain (designated CL). The thermal stability of the *S. pyogenes* V-CL construct is 36–37 °C, and a small number of changes in the collagen sequence has little effect on the thermal stability (Seo et al. 2010; An et al. 2016a). However, Cereceres et al. (2015) introduced 22 sequence replacements within the CL domain, often replacing Lys by Arg, which significantly increased the thermal stability and decreased the solubility. As discussed below, it is possible to make small changes in the CL collagen domain through site-directed mutagenesis or through the insertion (or replacement) of larger human sequences. The introduction of biologically- active human sequences within the bacterial collagenous domain is particularly valuable in helping to unravel and to develop new insights into structure-function relationships (Fig. 18.2). Although this Chapter is focused on the functional manipulations of the collagen structure, the bacterial collagen system has also been used to define the effect of Gly missense mutations on structure and function (Cheng et al. 2011; Yigit et al. 2016), and to explore the consequences of interruptions in the Gly-Xaa-Yaa repeat, which are present in non-fibrillar collagens (Hwang et al. 2010).

Examples of the use of recombinant collagen constructs to understand the relationship between collagen amino acid sequence and structure and function are described in the following sections.

18.4   Recombinant Collagens with Repeating Triple-Helix Segments

Repetitive elements are an underlying feature of collagens. The amino acid sequence consists of a linear series of Gly-Xaa-Yaa tripeptides. The structure of the collagen rod-like molecule is a supercoiled triple-helix, with a screw symmetry relating one chain to the next. The higher order fibril structure has a D = 67 nm axial repeat, generated by 67 nm axial staggering of triple-helical molecules. In addition, the charge and hydrophobic residues within the collagen molecule occur at regular intervals related to the D period (Hulmes et al. 1973). A number of reports have focused on manipulating the repeating elements in the triple-helix, and several of these are described in this section.

One of the earliest recombinant collagen studies was the expression of the human homotrimeric type II collagen molecule in a stably transfected tumour cell line HT1080 (Fertala et al. 1994), and recombinant type II molecules were shown to be able to form stable triple-helices and D-periodic fibrils (Fertala et al. 1996). A cassette model based on the type II collagen construct was used to investigate the stability and functional properties of collagen segments related to the D-period of collagen fibrils. Each type II collagen molecule consists of 4.4 D periods: D1, D2, D3, and D4, each 234 residues in length, followed by the shorter overlap segment D0.4 (Fig. 18.3). A strategy based on the deletion of individual D segments was developed to map the binding sites on fibrillar collagens for the DDR cell receptor and for chondrocyte binding (Arnold et al. 1997, 1998; Fertala et al. 2001; Majsterek et al. 2003; Leitinger et al. 2004). Such deletion studies suggested the D4 period is critical for adhesion, spreading and migration of chondrocytes (Fertala et al. 2001), and showed that only constructs containing the D1, D4 and the 0.4 D terminal sequence were able to form fibrils with a clear periodic banding pattern, indicating the importance of the N- and C-terminal segments (Steplewski et al. 2007). Based on this information, the Fertala group designed a novel recombinant collagen molecule which contained a series of three tandem D4 repeats (mD4; Fig. 18.3), and showed that scaffolds coated with this engineered collagen had properties superior to those obtained with original type II molecules (Ito et al. 2006).

The potential to produce a highly bioactive construct for biomedical applications led to smaller segments that include a functional domain that was assembled into a linear, contiguous trimeric structure (Peng et al. 2009). An extended segment of 15 triplets of type III collagen was used that contained an anti-type III collagen epitope that is proximal to an integrin-binding site. Expression was in Saccharomyces, with included co-expressed prolyl hydroxylase, with the trimer product being secreted 1 day after induction. The extent of Hyp formation was poor (Peng et al. 2009).
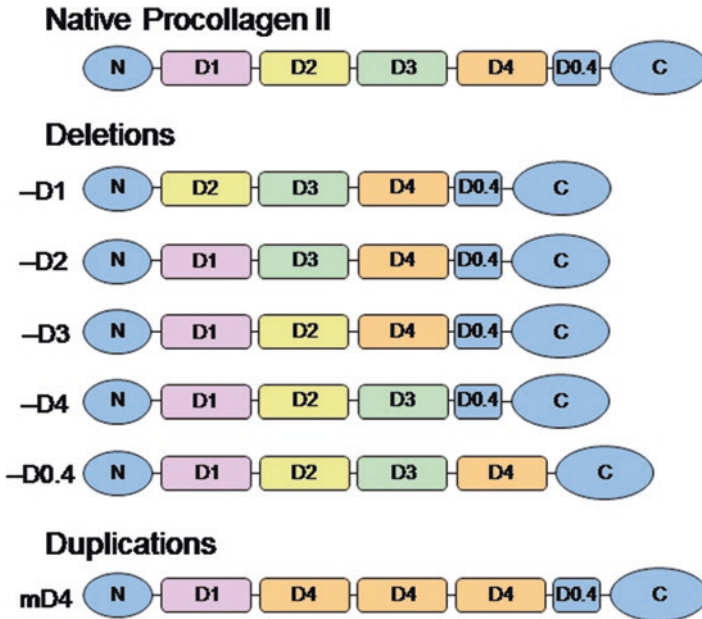
Fig. 18.3 Diagram of native type II procollagen (*top*), showing its D periodic segments (D1, D2, D3, D4 and D0.4), together with recombinant constructs generated in the Fertala laboratory that delete one segment (−D1, −D2, −D3, −D4) or include multiple repeats of the D4 segment (mD4; *bottom*) (Fertala et al. 2001; Ito et al. 2006)

Studies on a *de novo* designed recombinant collagen expressed in *E. coli* showed the importance of repeating patterns within the triple-helix in generating axially-periodic fibrils (Kaur et al. 2015). The collagen stability calculator (Persikov et al. 2005b; http://compbio.cs.princeton.edu/csc/) was used to design a 36 triplet segment including type I collagen sequences with a high triple-helix-forming propensity. This 36 triplet sequence was repeated three times, separated by (Gly-Pro-Pro)$_4$ sequences in the final recombinant construct. The foldon sequence was included at the C-terminal end to support trimerization, but the foldon was subsequently proteolytically removed from the purified protein. Although the construct was expressed in *E. coli* and hence had no Hyp present, it nevertheless formed a stable triple-helix and could aggregate into fibrillar structures with an axial periodicity of around 35 nm. This suggested a specific alternating gap/overlap structure had been formed with a repeat characterised by a single 36 triplet repeat and assembled possibly through electrostatic interactions (Kaur et al. 2015).

Recombinant *S. pyogenes* collagen-like Scl2 protein, with its (Gly-Xaa-Yaa)$_{79}$ CL collagenous domain, has been duplicated to make longer triple-helices and also dissected into small subsections. Since type I collagen has a triple-helix domain of (Gly-Xaa-Yaa)$_{338}$, attempts were made to reach this length by concatenating multiple bacterial collagen triple-helical domains, giving dimers (Yoshizumi et al. 2009), trimers and tetramers (Peng et al. 2014a). The thermal stability of the polymeric

forms did not significantly change with the increasing length. The utility of this approach to making larger molecules was limited as the production yield decreased noticeably as length increased (Peng et al. 2014a). Other duplications and manipulations of *S. pyogenes* collagenous proteins showed a length dependence of the thermal stability until one reached a minimum length of about 70–80 triplets, after which the stability was unaffected by further additional triplets (Humtsoe et al. 2005; Han et al. 2006). As well as joining bacterial collagens into larger structures, it is also possible to dissect them into smaller elements (Yu et al. 2011). The *S. pyogenes* Scl2 collagen domain CL was dissected into three almost equal size segments of 25–27 triplets, termed A, B and C from the N-terminal, which were expressed alone (A, B and C), together with the V-domain (V-A, V-B, and V-C), and as dimers or trimers (V-AA, V-AAA; V-BB, V-BBB). The individual segments were less stable than the intact CL domain, and repeats which increased triple-helical length led to an increased stability for the A and B segments; C dimers and trimers were insoluble and could not be studied (Yu et al. 2011). The length-dependent increase in stability appeared to plateau when around 55 or more triplets were present, although the observed stability was sequence-dependent. The central B segment is the most stable and a BBB trimer shows stability comparable to the native ABC trimer, while AAA trimer is less stable. Thus, the overall ABC stability appears to be determined by the more stable domain (Yu et al. 2011), consistent with the previous observation that was made for mammalian collagen (Steplewski et al. 2004).

18.5   Recombinant Collagens: Binding to Cell Receptors and ECM Proteins

In addition to its structural role in the extracellular matrix, collagen interacts with cell surface receptors, regulating cellular processes such as cell adhesion, proliferation and migration (Fig. 18.4). Collagen receptors include four members of the integrin family ($\alpha_2\beta_1$, $\alpha_1\beta_1$, $\alpha_{10}\beta_1$, $\alpha_{11}\beta_1$), as well as two Discoidin Domain Receptors (DDR1 and DDR2) and three members of the immunoglobulin superfamily of receptors (GPVI, LAIR1, OSCAR) (Leitinger and Hohenester 2007; An et al. 2016b). A triple-helical conformation is required for binding to these receptors. Original binding and functional studies were carried out using extracted collagens and collagen fragments, and subsequent peptide studies defined the Gly-Xaa-Yaa sequence required for receptor binding (Hamaia and Farndale 2014; An et al. 2016b). Such studies identified GFOGER, found in type I and II collagens, as the strongest integrin binding sequence, and GVMGFO as the strongest binding sequence for DDR receptors (Konitsiotis et al. 2008; Hamaia and Farndale 2014). When the sequence is known for a defined binding site in human collagen, it is possible to manipulate that sequence within a recombinant human collagen or to introduce that sequence into the triple-helix domain of a bacterial collagen and obtain the expected biological activity (Fig. 18.4). The manipulation of such sequences is
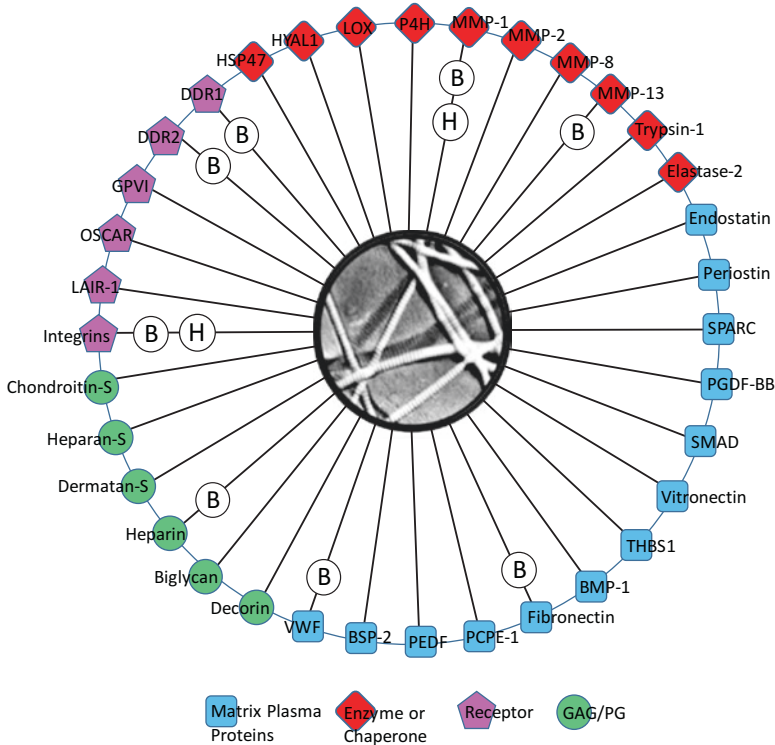
Fig. 18.4 A diagram of major categories of interactions of collagen with different proteins, based on An and Brodsky (2016). The interactions that have been manipulated within human recombinant collagen (H) or inserted within the bacterial collagen domain (B) are indicated next to the interacting component

expanding our understanding of relationships between sequence and structure and function, and allowing the creation of novel materials with targeted interactions.

Integrin binding has been manipulated in recombinant human type III collagen expressed in *S. cerevisiae*, in the laboratory of Szu-Wen Wang. This group developed a modular strategy for expressing and assembling full length type III collagen, using 12 fragments to cover the triple-helix domain (Chan et al. 2010); the α and β genes for human P4H were incorporated to catalyse Hyp formation. The recombinant type III collagen formed a stable triple-helix structure which supported adhesion of mammalian cells. Removal of five known integrin binding sites within the type III sequence (GROGER, GAOGER, GLOGEN, GLKGEN and GMOGER) abolished cell binding. After these original sites were removed, introduction of the strong type I integrin-binding sequence GFOGER restored cell binding, with the amount dependent on the number of GFOGER sites introduced and their location (Que et al. 2015). The potential to engineer integrin binding sites within different

types of human collagen presents opportunities for studying the principles of cell interactions as well as generating targeted biomaterials.

Advancement in our understanding of the interaction of collagen with integrin has also been made in the context of a bacterial collagen system. The Scl2 collagen-like protein from *S. pyogenes* does not bind to cells, but insertion of the human type I GFPGER motif within the collagen domain of the bacterial protein leads to integrin binding and cell adhesion (Seo et al. 2010, An et al. 2013; Peng et al. 2014b) (Fig. 18.2). The cell binding activity of these chimeric constructs confirms that Hyp is not necessary for integrin interactions, and bacterial collagen-like proteins containing GFPGER and GLPGER were shown to bind to both $\alpha_1\beta_1$ and $\alpha_2\beta_1$ integrin (Seo et al. 2010). In addition, a sequence GFPGEN computationally predicted to bind preferentially to the $\alpha1$ vs $\alpha2$ I domain was inserted, and the construct with this insertion showed the expected specificity (Seo et al. 2010). Hydrogels formed with cross-linked bacterial collagen containing GFPGER or GFPGEN showed selective adhesion to endothelial cells compared with smooth muscle cells, based on this differential binding (Cosgriff-Hernandez et al. 2010). Recent experiments on bacterial collagen constructs mapped the effect of replacing a single Gly within or nearby the GFPGER binding site on integrin binding, which may relate to collagen diseases (Yigit et al. 2016).

A recent report was published using bacterial collagen constructs to study binding of collagen to the DDR receptors (An et al. 2016b). The sequence GVMGFO has been shown to be essential for collagen binding to the DDR1 and DDR2 receptors (Konitsiotis et al. 2008). Insertion of a 6-triplet human type III collagen sequence containing this binding sequence within the bacterial collagen protein led to specific receptor binding, but the construct was unable to stimulate DDR auto-phosphorylation. This recombinant bacterial collagen, which can bind DDRs without inducing kinase activation, was shown to inhibit interactions between animal collagen and the DDRs both *in vitro* and in a cell migration assay.

In its structural and biological roles, collagen interacts with a wide range of molecules within the ECM and with serum proteins (Fig. 18.4) (An et al. 2016a). The interaction of collagen with fibronectin (Fn) is important for extracellular matrix assembly, and Fn binds very tightly to denatured collagen and less tightly to native collagen. The minimum sequence requirement for Fn binding was determined by inserting different lengths of type II collagen sequence around the proposed binding site into a recombinant bacterial collagen construct (An et al. 2014). Hyp in collagen was not essential for Fn binding, and denaturation of the recombinant collagen increased its Fn affinity. Cells cultured on bacterial collagen containing the Fn binding site demonstrated improved adhesion and differentiation, indicating the biological effectiveness of such chimeric recombinant bacterial collagens. Similar insertions of biologically active sequences have been reported for heparin and for vWF (Peng et al. 2014b; An et al. 2016b).

18.6   Recombinant Collagens: MMPs and Collagen Degradation

The degradation of collagen plays important biological roles during development, natural turnover and tissue repair. In addition, abnormal degradation of collagen is seen in osteoarthritis and in tumour progression. The collagen triple-helix is resistant to cleavage by most enzymes, and its degradation is mediated by a subset of matrix metalloproteinases (collagenases), MMP-1, MMP-8, MMP-13, and MT-MMP1, which all cleave fibrillar collagen molecules at one unique site about ¾ distance from the N-terminus. Many studies have employed model peptides to study MMP cleavage (Nagase and Fields 1996; Lauer-Fields and Fields 2002; Fields 2015), but there are some limitations in how accurately peptides model the MMP activity for native collagen cleavage. Transgenic mouse studies on a Mov13 mouse strain lacking a functional COL1A1 gene allowed transfection of a human α1(I) chain with specific amino acid substitutions around the known MMP-1 cleavage site, to clarify residues essential to cleavage (Liu et al. 1995). A major contribution of the transgenic mouse studies is the ability to examine the resultant physiological effects of these substitutions which prevented collagenase cleavage, such as excess collagen production in the skin and impaired post-partum involution of the uterus (Liu et al. 1995).

The expensive and time-consuming nature of transgenic mouse experiments has limited their applications, and alternative recombinant approaches have been pursued to enhance understanding of the MMP cleavage site in collagen. Changes in residues in the immediate vicinity of the collagenase cleavage site were introduced into recombinant human type III collagen expressed in *P. pastoris* yeast system, where P4H genes are also included (Williams and Olsen 2009). Williams and Olsen generated full length type III recombinant collagen, which formed a triple-helix with the same thermal stability and Hyp content seen in animal type III collagen. In addition, this recombinant type III collagen could form typical collagen fibrils. Replacing residues within or nearby the cleavage site by Pro residues decreased or eliminated collagenase activity, although it appears that the binding K_d was not affected.

The number of tripeptide units required for collagenase cleavage in type III was determined using the recombinant bacterial collagen system (Fig. 18.2) (Yu et al. 2012). The original *S. pyogenes* collagen domain could not be cleaved by MMPs. Introduction of triplets including and surrounding the human type III cleavage site within the bacterial CL collagen domain led to cleavage at the natural Gly-Ile site. The number of triplets required surrounding the scissile Gly-Ile-containing triplet was determined. When the human sequence included at least one triplet N-terminal and three triplets C-terminal to the Gly-Ile cleavage site, the protein was cleaved by MMP-1 and MMP-13. If fewer triplets were introduced, no cleavage was observed, indicating this is the minimum requirement for MMP recognition and cleavage. Interestingly, there is no Hyp in the collagen-like protein produced in this system, indicating the Hyp is not essential for collagenase cleavage. The bacterial system can also be used to create homotrimers of the α1(I) chain of type I collagen, and

homotrimers of the α2(I) chain of type I collagen, to assess their susceptibility to MMPs, or to vary the sequences around the cleavage site to further define the specificity (An, B and Brodsky, B, personal communication).

Although the studies reported here relate to fundamental knowledge about MMP-collagen interactions, the ability to introduce a collagenase cleavage site into a recombinant collagen material presents the opportunity for controlled degradation of the material. Manipulation of the sequences introduced or the number of cleavage sequences, or the specificity for different MMPs could allow for precise regulation of turnover of biomaterials created from such recombinant collagen.

18.7   Recombinant Collagen Fusion Proteins

From around 1998, the concept of making collagen-based hybrid proteins that included a bioactive component was envisaged in various patents (Gruskin et al. 1998). Later, hybrid constructs based on domains of different collagens were also suggested (Ramshaw and Werkmeister 2010). For example, hybrid constructs with bioactive segments for bone morphogenic protein, transforming growth factor β1 and decorin were proposed, with the collagen, either as an intact chain or as fragments, present to provide aggregation. Hybrid constructs of collagen with epidermal growth factor (EGF) and interleukin-2 (IL-2) were experimentally generated by Hayashi et al. (Hayashi et al. 2001, 2002). A fusion protein was created by using restriction enzymes to attach human mature EGF to the N-terminus of the triple-helical region of the human α1(III) chain, and the recombinant protein was expressed in a baculovirus system (Fig. 18.5). The triple-helical conformation was shown to be retained and there was mitogenic activity of the incorporated EGF component, both in solution, when immobilised on tissue culture plastic, and when included in a collagen gel (Hayashi et al. 2001). A fusion protein of type III collagen with IL-2 attached at the N-terminus was made using a similar strategy. This material demonstrated biological activity, where the baculovirus cell supernatant showed mitogenic activity similar to that seen for non-conjugated IL-2 (Hayashi et al. 2002). In both
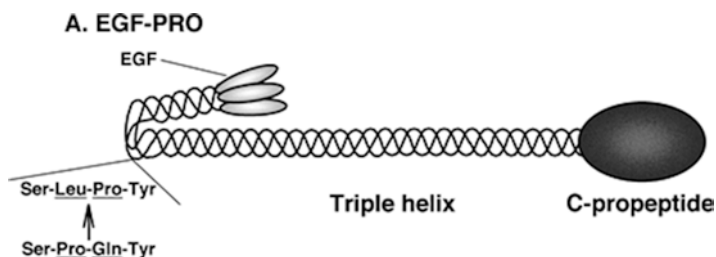
Fig. 18.5 Schematic diagram of a hybrid construct of collagen fused with epidermal growth factor (EGF) expressed in a baculovirus system (Reprinted with permission from Hayashi et al. 2001)

cases, the designed constructs could have potential clinical applications, in wound healing and in cancer therapy, respectively (Hayashi et al. 2001, 2002).

Trimerization domains adjacent to collagenous domains have been manipulated in the recombinant bacterial collagen system as well (Fig. 18.2). In one set of constructs, the original V bacterial trimerization domain is deleted and is replaced by an engineered three-stranded coiled-coil trimerization domain added at either the N- or C-terminus, or at both ends of the *S. pyogenes* Scl2 bacterial collagen domain (Yoshizumi et al. 2011). The highly stable coiled-coil domain (T_m ~ 90 °C) provided an anchor for the collagenous domain (T_m ~37 °C), and refolding kinetics of the collagen were comparable to the original protein with the V-domain (Yoshizumi et al. 2011). The foldon domain, a trimeric beta-propeller structure from fibritin protein in bacteriophage T4, has also been used as a trimerization domain for collagen sequences and for artificial designed $(Gly-Xaa-Yaa)_n$ sequences expressed in *E. coli* (Du et al. 2008; Kaur et al. 2015).

The fusion of a collagen domain to silk sequences has been used to generate recombinant collagen that can be incorporated into well characterized materials. Repetitive *Bombyx mori* silk consensus sequences were added to the C-terminal end of the *S. pyogenes* Scl2 bacterial collagen domain, as a strategy for promoting specific binding of the collagen to solid state silk materials, which have well-defined mechanical and material properties (Fig. 18.6) (An et al. 2013). The silk-like sequence comprised three, six or nine repeats of $(Gly-Ala-Gly-Ala-Gly-Ser)_n$. The binding affinity to silk films and porous scaffolds was determined by the number of repeats, with the $(Gly-Ala-Gly-Ala-Gly-Ser)_6$ construct showing the best binding. The bacterial collagen domain was modified to include integrin-binding and fibronectin-binding sequences (An et al. 2013), and the binding of these collagen-silk constructs to silk biomaterials enhanced mesenchymal stem cell spreading and proliferation (Fig. 18.6). This system illustrates a method whereby silk films and materials, which typically show excellent stability and mechanical properties, can be readily modified to include one or more designed collagen functions that enhance the biological activity for tissue engineering applications.

18.8   Applications for Designed Recombinant Collagens

One major impetus for the development of recombinant collagen production was the desire to generate materials for biomedical applications. An important consideration for this goal is the ability to achieve large scale production and easy purification of the recombinant proteins. Potentially useful systems include yeasts, especially *Pichia* where yields in excess of 1.5 g/L for collagen and 14 g/L for gelatin have been achieved (Yang et al. 2004; Báez et al. 2005), and transgenic animals, where yields of 10 g/L in milk have been reported (Toman et al. 1999). More recently plant systems are showing promise (Stein et al. 2009; Xu et al. 2011), but a significant investment is needed to optimise and/or breed suitable strains. *E. coli* can readily provide excellent yields (Peng et al. 2012), and can be used to express collagenous
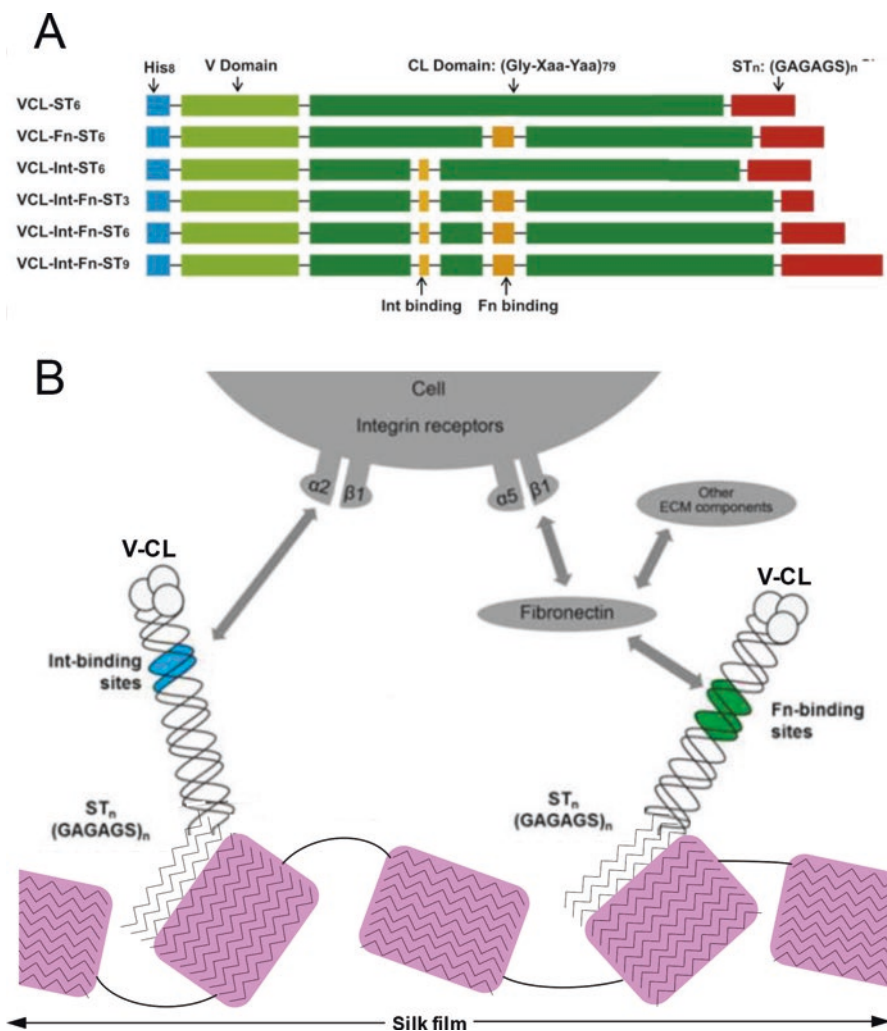
Fig. 18.6 Schematic diagram of bacterial collagen constructs containing silk sequences as well as integrin- and/or fibronectin-binding domains (*top*). The inclusion of a fibronecting-binding domain is shown to increase cell proliferation (*middle panel*). An illustration of the directed interactions with cells and silk films is shown below (Based on An et al. 2013)

proteins which do not require Hyp for stability, such as bacterial collagens. For expression of animal collagens in *E. coli*, it will be necessary to further develop the efficiency of inserted P4H activity (Pinkas et al. 2011), although success with some collagen fragments was reported recently (Tang et al. 2016). In recent years, progress has been made in developing recombinant animal collagens and bacterial collagens for potential products, and this progress will be reviewed here.

18.8.1 Applications of Recombinant Animal Collagens

The *Pichia* system has been used successfully to express recombinant human type I, II and type III collagens, as well as a number of other collagen genetic types, and recombinant human type I and III collagens are now commercially available. For instance, in the commercially available type I collagen, human α1(I) and α2(I) chains were co-expressed with the α- and β-subunits of P4H in *P. pastoris*, followed by a controlled proteinase digestion to obtain the triple-helix domain. Compared with collagens extracted from animal tissues, the recombinant human collagens have a number of important advantages (for reviews, see Bulleid et al. 2000; Olsen et al. 2003; Báez et al. 2005; Ruggiero and Koch 2008; Werkmeister and Ramshaw 2012; Ramshaw et al. 2014). The recombinant collagen provides a defined material, which is biochemically pure, with none of the batch-to-batch variation possible for extracted collagens. In addition, these collagens are virus- and disease-free, in contrast to animal-derived collagens where there are risks of transmission of infectious agents. Recombinant production also provides a route to obtain minor collagens that are of low natural abundance and are not readily prepared from tissue by traditional protein chemistry techniques. The human collagen sequence is the one most commonly expressed as a recombinant protein, while all extracted collagens will come from animals (usually bovine but also porcine). Expressed human collagens should lack immunogenicity, while retaining the biological interactions of natural collagens, including cell-matrix interactions. On the other hand, recombinant materials do not recreate all of the features of native collagen. Although they can be manipulated to introduce hydroxylation of Pro, they lack other post-translational modifications such as hydroxylation and glycosylation of Lys and lysyl oxidase-mediated cross-linking. As a result, the recombinant collagen will not necessarily be suitable for all biomedical applications. They are useful for production, for example, of gels and sponges, but are not suitable to replace devices made from stabilised intact tissues, such as heart valves, where the natural strength and durability probably cannot be achieved, even with extensive cross-linking. In addition, commercially available recombinant collagens are still expensive and the amounts available limited.

A number of groups are exploring potential applications for the available recombinant human collagen proteins. As recombinant collagens are not as readily available in quantity as tissue-derived collagens, certain applications which use less material and require very high purity are more attractive targets. The extensive studies on corneal replacements demonstrate well this target selection (Fagerholm et al. 2014) (Fig. 18.7). Recombinant human type III, as well as type I collagen, have been used to generate corneal replacements (Liu et al. 2008). It was found that 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide-based crosslinking of 13.7 % recombinant collagen type I and type III gels gave optically-transparent devices that have been investigated as corneal replacements, with the type III collagen-based device showing better mechanical properties (Liu et al. 2008). Further, these devices

rh-Collagen 3

Phosphocholine
derivative

Composite gel

Corneal
replacement
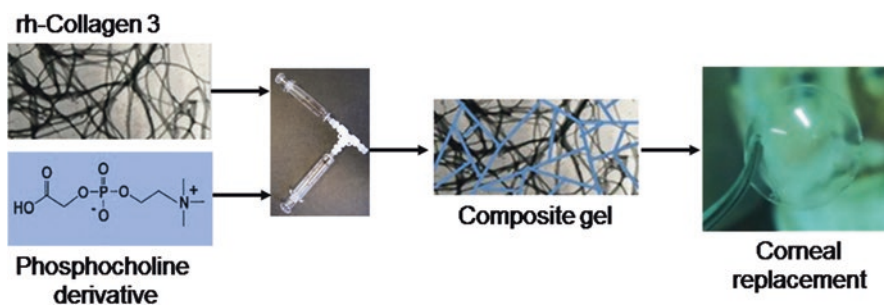
Fig. 18.7 Construction of corneal replacements from recombinant human type III collagen and a phosphocholine derivative (Reprinted with permission from Buznyk et al. 2015)

have shown good performance at 4 years, indicating that they could provide a potentially safe alternative to donor transplantation (Fagerholm et al. 2014). Other examples of projects utilizing recombinant collagens include investigation of recombinant human type II collagen gels, in conjunction with chondrocytes, for cartilage repair (Pulkkinen et al. 2013); use of recombinant human type I and III collagen sponges for the culture of gingival fibroblasts as potential gingival repair materials (Yamada et al. 2006); and evaluation of recombinant type I collagen in combination with BMP-2 as a bone graft material (Yang et al. 2004). In addition, the use of the recombinant collagens as a source of gelatin has been explored (Olsen et al. 2005).

One distinctive and critical feature of recombinant collagens compared with extracted collagens is the potential for modifying the amino acid sequence to intentionally affect interactions and degradation desired for targeted applications. It has been proposed that rational engineering of collagens can generate scaffolds in tissue engineering and products to control fibrosis and accelerate wound healing (John et al. 1999; Fertala et al. 2001; Majsterek et al. 2003; Ito et al. 2006). This field is in its early stages, but several groups are moving in this direction. Manipulation of the D segments within a full length type II collagen molecule (see Section 18.4 above) indicated that the D4 segment is very important in cell spreading and mobility. A modified full length construct containing three tandem D4 segments was incorporated into a scaffold for chondrocyte growth (Fig. 18.3) and shown to create cartilaginous material superior to that formed on a native type II collagen scaffold (Ito et al. 2006). The human type III collagen constructs where the integrin binding sites have been manipulated also hold promise for tissue engineering (Que et al. 2015). The specific amino acid sequences required for an increasing number of interaction sites for collagen have been identified (Sweeney et al. 2008; An et al. 2016a) (Fig. 18.4), and these defined interacting sequences will provide more opportunities for directed engineering of native collagen sequences for tissue engineering, drug delivery and other biomedical applications.

18.8.2 Applications of Bacterial Collagens

In addition to helping to unravel structure-function relationships, constructs of bacterial collagen containing biologically-active sequences from human collagen have potential value for the generation of bioengineered materials (Fig. 18.2). Products based on bacterial collagens expressed in *E. coli* would have strong advantages. The recombinant bacterial proteins are non-cytotoxic and non-immunogenic (Peng et al. 2010). Their bacterial origin is consistent with the trend towards non-animal products for general compliance strategies or for products where cultural or lifestyle choices may drive demand. Importantly, the bacterial collagen proteins are fully stable without the need for secondary modification to introduce Hyp, circumventing the need for introducing P4H genes. The few cases where Hyp is a mandatory residue for effective function may be resolved either through modification of the medium to give Hyp incorporation (Gruskin et al. 1998; Buechter et al. 2003) or utilisation of a viral hydroxylase enzyme (Luther et al. 2011; Rutschmann et al. 2014) or adaptation of the pathways (Pinkas et al. 2011). The bacterial collagen proteins can be readily produced in *E. coli* by fermentation in bioreactors with yields of up to 19 g/L (Peng et al. 2012). The initial product is normally V-CL, with the attached non-triple helical V trimerization domain, but large scale purification through precipitation and proteolysis removes the V domain, yielding only triple-helical material (Peng et al. 2014c). As discussed above, it has proved to be straight-forward to insert biologically-active sequences from human collagen within the bacterial collagen domain which can tune a material for a specific environment and application. There are also limitations with the recombinant bacterial collagens. Unlike interstitial collagens, the bacterial collagens do not readily form large fibrous aggregates (Yoshizumi et al. 2009), requiring different strategies for fabrication. However, the diversity of collagen-based medical devices is large (Ramshaw et al. 2009) and in many cases, especially those that use soluble purified collagens, bacterial collagens could provide potential alternatives. Products will not have the strength and durability seen in modified natural tissues, but could be very suitable for many applications, such as active coatings. One additional concern for recombinant collagens expressed in *E. coli* is the potential for endotoxin contamination, although a variety of methods have been described for reducing endotoxin content (Zhang et al. 2014).

The ease of manipulating biologically-active sequences and adding domains to provide physical advantages makes the bacterial collagen system attractive for targeted applications (Fig. 18.2). A range of opportunities exist to incorporate bacterial collagens into scaffolds in the form of porous sponges, films, gels or filaments, all of which could incorporate cells, growth factors or drugs for control of tissue regeneration. A number of published reports illustrate ways in which bacterial collagen may be modified and adapted for outstanding biomedical needs. As described above (Section 18.5), the Scl2 collagenous protein of *S. pyogenes* contains no intrinsic biological activity, but insertion of the GFPGER sequence, which binds $\alpha_1\beta_1$ and

$\alpha_2\beta_1$ integrin, leads to cell adhesion and migration (Cosgriff-Hernandez et al. 2010). The collagen molecules can be fabricated into a functional PEG-polymeric composite material, using a photo-polymerisation strategy. The same cell specificity that had been observed for the individual designed molecules was retained in the composite material (Cosgriff-Hernandez et al. 2010). This system was optimized to have maximal endothelial cell attachment and spreading by increasing the substrate modulus and lowering the functionalization density, while increasing protein concentration (Browning et al. 2014). The hydrogel alone does not have suitable mechanical properties and durability for a long-term vascular implant, but this has been addressed by development of composite materials whereby new or existing vascular materials, for example polyurethanes, can be adapted into a multilayer device through incorporation of the functional hydrogel at the blood contact interface (Browning et al. 2012). In other applications, a hydrogel coating containing *S. pyogenes* bacterial collagen with the GFPGER sequence was shown to enhance osteoblast adhesion and spreading on titanium materials (Bronk et al. 2014). The system was also modified to improve collagen stability and create hydrated microspheres for wound healing (Cereceres et al. 2015). In addition to biological properties directed by inserted sequences, bacterial collagens were found to have a number of intrinsic properties useful in biomedical devices, such as being non-thrombogenic (Browning et al. 2012), inhibiting colonisation of certain bacteria and preventing biofilm formation (Bronk et al. 2014).

A different hydrogel concept has been examined for use in cartilage repair (Parmar et al. 2015). In this instance, the bacterial collagen was modified by attaching specific hyaluronic acid- and chondroitin sulphate-binding peptides via a heterobifunctional NHS-PEG-acrylate spacer that was reacted with a Cys residue at the N-terminal end of each peptide, while biodegradable crosslinking between collagen molecules was achieved using the acrylate-modified collagen and an MMP7-sensitive peptide with Cys residues at both N- and C-terminals. Additional crosslinking is obtained by interaction of the binding peptides with hyaluronic acid and chondroitin sulphate. These could be added to the system during fabrication, but of more interest is when they are derived from cellular biosynthesis in association with the implanted material (Parmar et al. 2015). Initial *in vitro* studies using human mesenchymal stem cells indicated that addition of the binding peptides enhanced the stem cell viability. Further manipulation of the proportions of the binding and cross-linking peptides should allow optimisation of the mechanical properties and the kinetics of chondrogenesis and matrix accumulation, while balancing cell mediated hydrogel degradation through cellular MMP7 production (Parmar et al. 2015).

18.9 Summary and Future Directions

Significant progress has been made in expressing recombinant human collagens, and using this material for specific appropriate applications, such as the production of bioengineered corneal replacements. In addition, the recombinant bacterial collagen system allows for ease of manipulation of sequences, so that materials with a specific subset of biological signals, rather than all the signals present in human collagen, can be developed to selectively promote cell adhesion, cleavage or other desired processes. Progress has been made in the expression and manipulation of biological signals in recombinant collagens, which forms the basis for creative applications of these systems to meet outstanding biomaterial and biomedical needs.

References

Abraham LC, Zuena E, Perez-Ramirez B, Kaplan DL (2008) Guide to collagen characterization for biomaterial studies. J Biomed Mater Res B Appl Biomater 87:264–285

Ackerman MS, Bhate M, Shenoy N, Beck K, Ramshaw JAM, Brodsky B (1999) Sequence dependence of the folding of collagen-like peptides. Single amino acids affect the rate of triple-helix nucleation. J Biol Chem 274:7668–7673

An B, Brodsky B (2016) Collagen binding to OSCAR: the odd couple. Blood 127:521–522

An B, DesRochers TM, Qin G, Xia X, Thiagarajan G, Brodsky B, Kaplan DL (2013) The influence of specific binding of collagen-silk chimeras to silk biomaterials on hMSC behavior. Biomaterials 34:402–412

An B, Abbonante V, Yigit S, Balduini A, Kaplan DL, Brodsky B (2014) Definition of the native and denatured type II collagen binding site for fibronectin using a recombinant collagen system. J Biol Chem 289:4941–4951

An B, Lin YS, Brodsky B (2016a) Collagen interactions: drug design and delivery. Adv Drug Deliv Rev 97:69–84

An B, Abbonante V, Xu H, Gavriilidou D, Yoshizumi A, Bihan D, Farndale RW, Kaplan DL, Balduini A, Leitinger B, Brodsky B (2016b) Recombinant collagen engineered to bind to discoidin domain receptor functions as a receptor inhibitor. J Biol Chem 291:4343–4355

Arnold WV, Sieron AL, Fertala A, Bächinger HP, Mechling D, Prockop DJ (1997) A cDNA cassette system for the synthesis of recombinant procollagens. Variants of procollagen II lacking a D-period are secreted as triple-helical monomers. Matrix Biol 16:105–116

Arnold WV, Fertala A, Sieron AL, Hattori H, Mechling D, Bächinger HP, Prockop DJ (1998) Recombinant procollagen II: deletion of D period segments identifies sequences that are required for helix stabilization and generates a temperature-sensitive N-proteinase cleavage site. J Biol Chem 273:31822–31828

Báez J, Olsen D, Polarek JW (2005) Recombinant microbial systems for the production of human collagen and gelatin. Appl Microbiol Biotechnol 69:245–252

Bella J, Eaton M, Brodsky B, Berman HM (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science 266:75–81

Bella J, Brodsky B, Berman HM (1995) Hydration structure of a collagen peptide. Structure 3:893–906

Boudko SP, Engel J, Bächinger HP (2012) The crucial role of trimerization domains in collagen folding. Int J Biochem Cell Biol 44:21–32

Boydston JA, Chen P, Steichen CT, Turnbough CL (2005) Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis*. J Bacteriol 187:5310–5317

Brodsky B, Ramshaw JAM (1997) The collagen triple-helix structure. Matrix Biol 15:545–554

Brodsky B, Persikov AV (2005) Molecular structure of the collagen triple helix. Adv Protein Chem 70:301–339

Brodsky B, Thiagarajan G, Madhan B, Kar K (2008) Triple-helical peptides: an approach to collagen conformation, stability, and self-association. Biopolymers 89:345–353

Bronk JK, Russell BH, Rivera JJ, Pasqualini R, Arap W, Höök M, Barbu EM (2014) A multifunctional streptococcal collagen-mimetic protein coating prevents bacterial adhesion and promotes osteoid formation on titanium. Acta Biomater 10:3354–3362

Browning MB, Dempsey D, Guiza V, Becerra S, Rivera J, Russell B, Höök M, Clubb F, Miller M, Fossum T, Dong JF, Bergeron AL, Hahn M, Cosgriff-Hernandez E (2012) Multilayer vascular grafts based on collagen-mimetic proteins. Acta Biomater 8:1010–1021

Browning MB, Guiza V, Russell B, Rivera J, Cereceres S, Höök M, Hahn MS, Cosgriff-Hernandez EM (2014) Endothelial cell response to chemical, biological, and physical cues in bioactive hydrogels. Tissue Eng Part A 20:3130–3141

Buechter DD, Paolella DN, Leslie BS, Brown MS, Mehos KA, Gruskin EA (2003) Co-translational incorporation of trans-4-hydroxyproline into recombinant proteins in bacteria. J Biol Chem 278:645–650

Buevich A, Baum J (2001) Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases. Philos Trans R Soc Lond Ser B Biol Sci 356:159–168

Bulleid NJ, John DC, Kadler KE (2000) Recombinant expression systems for the production of collagen. Biochem Soc Trans 28:350–353

Buznyk O, Pasyechnikova N, Islam MM, Iakymenko S, Fagerholm P, Griffith M (2015) Bioengineered corneas grafted as alternatives to human donor corneas in three high-risk patients. Clin Transl Sci 8:558–562

Byers PH, Cole WG (2002) Osteogenesis imperfecta. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders: molecular, genetic and medical aspects, 2nd edn. Wiley-Liss, New York, pp 385–430

Cereceres S, Touchet T, Browning MB, Smith C, Rivera J, Höök M, Whitfield-Cargile C, Russell B, Cosgriff-Hernandez E (2015) Chronic wound dressings based on collagen-mimetic proteins. Adv Wound Care 4:444–456

Chan SW, Hung SP, Raman SK, Hatfield GW, Lathrop RH, Da Silva NA, Wang SW (2010) Recombinant human collagen and biomimetic variants using a de novo gene optimized for modular assembly. Biomacromolecules 11:1460–1469

Cheng H, Rashid S, Yu Z, Yoshizumi A, Hwang E, Brodsky B (2011) Location of glycine mutations within a bacterial collagen protein affects degree of disruption of triple-helix folding and conformation. J Biol Chem 286:2041–2046

Cosgriff-Hernandez E, Hahn MS, Russell B, Wilems T, Munoz-Pinto D, Browning MB, Rivera J, Höök M (2010) Bioactive hydrogels based on designer collagens. Acta Biomater 6:3969–3977

Ding S, Pinkas DM, Barron AE (2012) Synthesis and assembly of functional high molecular weight adiponectin multimers in an engineered strain of *Escherichia coli*. Biomacromolecules 13:1035–1042

Du C, Wang M, Liu J, Pan M, Cai Y, Yao J (2008) Improvement of thermostability of recombinant collagen-like protein by incorporating a foldon sequence. Appl Microbiol Biotechnol 79:195–202

Fagerholm P, Lagali NS, Ong JA, Merrett K, Jackson WB, Polarek JW, Suuronen EJ, Liu Y, Brunette I, Griffith M (2014) Stable corneal regeneration four years after implantation of a cell-free recombinant human collagen scaffold. Biomaterials 35:2420–2427

Fertala A, Sieron AL, Ganguly A, Li SW, Ala-Kokko L, Anumula KR, Prockop DJ (1994) Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080). Biochem J 298:31–37

Fertala A, Holmes DF, Kadler KE, Sieron AL, Prockop DJ (1996) Assembly in vitro of thin and thick fibrils of collagen II from recombinant procollagen II. The monomers in the tips of thick fibrils have the opposite orientation from monomers in the growing tips of collagen I fibrils. J Biol Chem 271:14864–14869

Fertala A, Han WB, Ko FK (2001) Mapping critical sites in collagen II for rational design of gene-engineered proteins for cell-supporting materials. J Biomed Mater Res 57:48–58

Fields GB (2015) New strategies for targeting matrix metalloproteinases. Matrix Biol 44-46:239–246

Fowler SJ, Jose S, Zhang X, Deutzmann R, Sarras MP, Boot-Handford RP (2000) Characterization of hydra type IV collagen. Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. J Biol Chem 275:39589–39599

Francois J, Herbage D, Junqua S (1980) Cockroach collagen: isolation, biochemical and biophysical characterization. Eur J Biochem 112:389–396

Frischholz S, Beier F, Girkontaite I, Wagner K, Pöschl E, Turnay J, Mayer U, von der Mark K (1998) Characterization of human type X procollagen and its NC-1 domain expressed as recombinant proteins in HEK293 cells. J Biol Chem 273:4547–4555

Fukuda K, Hori H, Utani A, Burbelo PD, Yamada Y (1997) Formation of recombinant triple-helical [α1(IV)]₂ α2(IV) collagen molecules in CHO cells. Biochem Biophys Res Commun 231:178–182

Geddis AE, Prockop DJ (1993) Expression of human COL1A1 gene in stably transfected HT1080 cells: the production of a thermostable homotrimer of type I collagen in a recombinant system. Matrix 13:399–405

Ghosh N, McKillop TJ, Jowitt TA, Howard M, Davies H, Holmes DF, Roberts IS, Bella J (2012) Collagen-like proteins in pathogenic *E. coli* strains. PLoS One 7:e37872

Gruskin EA, Buechter DD, Zhang G, Connolly K (1998) Amino acid modified polypeptides. US Patent 5821089-A

Hamaia S, Farndale RW (2014) Integrin recognition motifs in the human collagens. Adv Exp Med Biol 819:127–142

Han R, Zwiefka A, Caswell CC, Xu Y, Keene DR, Lukomska E, Zhao Z, Höök M, Lukomski S (2006) Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. Appl Microbiol Biotechnol 72:109–115

Hashizume F, Hino S, Kakehashi M, Okajima T, Nadano D, Aoki N, Matsuda T (2008) Development and evaluation of transgenic rice seeds accumulating a type II-collagen tolerogenic peptide. Transgenic Res 17:1117–1129

Hayashi M, Tomita M, Yoshizato K (2001) Production of EGF-collagen chimeric protein which shows the mitogenic activity. Biochim Biophys Acta 1528:187–195

Hayashi M, Tomita M, Yoshizato K (2002) Interleukin-2-collagen chimeric protein which liberates interleukin-2 upon collagenolysis. Protein Eng 15:429–436

Hulmes DJ, Miller A, Parry DAD, Piez KA, Woodhead-Galloway J (1973) Analysis of the primary structure of collagen for the origins of molecular packing. J Mol Biol 79:137–148

Humtsoe JO, Kim JK, Xu Y, Keene DR, Höök M, Lukomski S, Wary KK (2005) A streptococcal collagen-like protein interacts with the α2β1 integrin and induces intracellular signaling. J Biol Chem 280:13848–13857

Hwang ES, Thiagarajan G, Parmar AS, Brodsky B (2010) Interruptions in the collagen repeating tripeptide pattern can promote supramolecular association. Protein Sci 19:1053–1064

Ito H, Steplewski A, Alabyeva T, Fertala A (2006) Testing the utility of rationally engineered recombinant collagen-like proteins for applications in tissue engineering. J Biomed Mater Res A 76:551–560

John DC, Watson R, Kind AJ, Scott AR, Kadler KE, Bulleid NJ (1999) Expression of an engineered form of recombinant procollagen in mouse milk. Nat Biotechnol 17:385–389

Kaur PJ, Strawn R, Bai H, Xu K, Ordas G, Matsui H, Xu Y (2015) The self-assembly of a mini-fibril with axial periodicity from a designed collagen-mimetic triple helix. J Biol Chem 290:9251–9261

Kersteen EA, Higgin JJ, Raines RT (2004) Production of human prolyl 4-hydroxylase in *Escherichia coli*. Protein Expr Purif 38:279–291

Konitsiotis AD, Raynal N, Bihan D, Hohenester E, Farndale RW, Leitinger B (2008) Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. J Biol Chem 283:6861–6868

Kramer RZ, Venugopal MG, Bella J, Mayville P, Brodsky B, Berman HM (2000) Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. J Mol Biol 301:1191–1205

Krejci E, Coussen F, Duval N, Chatel JM, Legay C, Puype M, Vandekerckhove J, Cartaud J, Bon S, Massoulié J (1991) Primary structure of a collagenic tail peptide of Torpedo acetylcholinesterase: co-expression with catalytic subunit induces the production of collagen-tailed forms in transfected cells. EMBO J 10:1285–1293

Kuivaniemi H, Tromp G, Prockop DJ (1991) Mutations in collagen genes: causes of rare and some common diseases in humans. FASEB J 5:2052–2060

Lauer-Fields JL, Fields GB (2002) Triple-helical peptide analysis of collagenolytic protease activity. Biol Chem 383:1095–1105

Leitinger B, Hohenester E (2007) Mammalian collagen receptors. Matrix Biol 26:146–155

Leitinger B, Steplewski A, Fertala A (2004) The D2 period of collagen II contains a specific binding site for the human discoidin domain receptor, DDR2. J Mol Biol 344:993–1003

Li HC, Huang CC, Chen SF, Chou MY (2005) Assembly of homotrimeric type XXI minicollagen by coexpression of prolyl 4-hydroxylase in stably transfected *Drosophila melanogaster* S2 cells. Biochem Biophys Res Commun 336:375–385

Liu X, Wu H, Byrne M, Jeffrey J, Krane S, Jaenisch R (1995) A targeted mutation at the known collagenase cleavage site in mouse type I collagen impairs tissue remodeling. J Cell Biol 130:227–237

Liu W, Merrett K, Griffith M, Fagerholm P, Dravida S, Heyne B, Scaiano JC, Watsky MA, Shinozaki N, Lagali N, Munger R, Li F (2008) Recombinant human collagen for tissue engineered corneal substitutes. Biomaterials 29:1147–1158

Luther KB, Hülsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J Biol Chem 286:43701–43709

Majsterek I, McAdams E, Adachi E, Dhume ST, Fertala A (2003) Prospects and limitations of the rational engineering of fibrillar collagens. Protein Sci 12:2063–2072

Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R, Ruggiero F, Theisen M (2002) Hydroxylated human homotrimeric collagen I in *Agrobacterium tumefaciens*-mediated transient expression and in transgenic tobacco plant. FEBS Lett 515:114–118

Myllyharju J, Nokelainen M, Vuorela A, Kivirikko KI (2000) Expression of recombinant human type I-III collagens in the yeast *Pichia pastoris*. Biochem Soc Trans 28:353–357

Nagase H, Fields GB (1996) Human matrix metalloproteinase specificity studies using collagen sequence-based synthetic peptides. Biopolymers 40:399–416

Nokelainen M, Helaakoski T, Myllyharju J, Notbohm H, Pihlajaniemi T, Fietzek PP, Kivirikko KI (1998) Expression and characterization of recombinant human type II collagens with low and high contents of hydroxylysine and its glycosylated forms. Matrix Biol 16:329–338

Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perälä M, Hämäläinen ER, Jarvinen M, Polarek J (2003) Recombinant collagen and gelatin for drug delivery. Adv Drug Deliv Rev 55:1547–1567

Olsen D, Jiang J, Chang R, Duffy R, Sakaguchi M, Leigh S, Lundgard R, Ju J, Buschman F, Truong-Le V, Pham B, Polarek JW (2005) Expression and characterization of a low molecular

weight recombinant human gelatin: development of a substitute for animal-derived gelatin with superior features. Protein Expr Purif 40:346–357

Ortega N, Werb Z (2002) New functional roles for non-collagenous domains of basement membrane collagens. J Cell Sci 115:4201–4214

Parmar PA, Chow LW, St-Pierre JP, Horejs CM, Peng YY, Werkmeister JA, Ramshaw JAM, Stevens MM (2015) Collagen-mimetic peptide-modifiable hydrogels for articular cartilage regeneration. Biomaterials 54:213–225

Peng YY, Werkmeister JA, Vaughan PR, Ramshaw JAM (2009) Constructs for the expression of repeating triple-helical protein domains. Biomed Mater 4:015006

Peng YY, Yoshizumi A, Danon SJ, Glattauer V, Prokopenko O, Mirochnitchenko O, Yu Z, Inouye M, Werkmeister JA, Brodsky B, Ramshaw JAM (2010) A *Streptococcus pyogenes* derived collagen-like protein as a non-cytotoxic and non-immunogenic cross-linkable biomaterial. Biomaterials 31:2755–2761

Peng YY, Howell L, Stoichevska V, Werkmeister JA, Dumsday GJ, Ramshaw JAM (2012) Towards scalable production of a collagen-like protein from *Streptococcus pyogenes* for biomedical applications. Microb Cell Factories 11:146

Peng YY, Stoichevska V, Howell L, Madsen S, Werkmeister JA, Dumsday GJ, Ramshaw JAM (2014a) Preparation and characterization of monomers to tetramers of a collagen-like domain from *Streptococcus pyogenes*. Bioengineered 5:378–385

Peng YY, Stoichevska V, Schacht K, Werkmeister JA, Ramshaw JAM (2014b) Engineering multiple biological functional motifs into a blank collagen-like protein template from *Streptococcus pyogenes*. J Biomed Mater Res A 102:2189–2196

Peng YY, Stoichevska V, Madsen S, Howell L, Dumsday GJ, Werkmeister JA, Ramshaw JAM (2014c) A simple cost-effective methodology for large-scale purification of recombinant non-animal collagens. Appl Microbiol Biotechnol 98:1807–1815

Perret S, Merle C, Bernocco S, Berland P, Garrone R, Hulmes DJ, Theisen M, Ruggiero F (2001) Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. J Biol Chem 276:43693–43698

Perret S, Eble JA, Siljander PR, Merle C, Farndale RW, Theisen M, Ruggiero F (2003) Prolyl hydroxylation of collagen type I is required for efficient binding to integrin α1β1 and platelet glycoprotein VI but not to α2β1. J Biol Chem 278:29873–29879

Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B (2000) Amino acid propensities for the collagen triple-helix. Biochemistry 39:14960–14967

Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B (2002) Peptide investigations of pairwise interactions in the collagen triple-helix. J Mol Biol 316:385–394

Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B (2005a) Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. Biochemistry 44:1414–1422

Persikov AV, Ramshaw JAM, Brodsky B (2005b) Prediction of collagen stability from amino acid sequence. J Biol Chem 280:19343–19349

Pihlajamaa T, Perälä M, Vuoristo MM, Nokelainen M, Bodo M, Schulthess T, Vuorio E, Timpl R, Engel J, Ala-Kokko L (1999) Characterization of recombinant human type IX collagen Association of alpha chains into homotrimeric and heterotrimeric molecules. J Biol Chem 274:22464–22468

Pinkas DM, Ding S, Raines RT, Barron AE (2011) Tunable, post-translational hydroxylation of collagen domains in *Escherichia coli*. ACS Chem Biol 6:320–324

Pulkkinen HJ, Tiitu V, Valonen P, Jurvelin JS, Rieppo L, Töyräs J, Silvast TS, Lammi MJ, Kiviranta I (2013) Repair of osteochondral defects with recombinant human type II collagen gel and autologous chondrocytes in rabbit. Osteoarthr Cartil 21:481–490

Que RA, Chan SW, Jabaiah AM, Lathrop RH, Da Silva NA, Wang SW (2015) Tuning cellular response by modular design of bioactive domains in collagen. Biomaterials 53:309–317

Ramachandran GN, Kartha G (1955) Structure of collagen. Nature 176:593–595

Ramshaw JAM, Peng YY, Glattauer V, Werkmeister JA (2009) Collagens as biomaterials. J Mater Sci Mater Med 20(Suppl 1):S3–S8

Ramshaw JAM, Werkmeister JA (2010) Novel collagen constructs. Patent Appl: WO2010071938-A1

Ramshaw JAM, Werkmeister JA, Dumsday GJ (2014) Bioengineered collagens: emerging directions for biomedical materials. Bioengineered 5:227–233

Rasmussen M, Jacobsson M, Björck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. J Biol Chem 278:32313–32316

Ricard-Blum S (2011) The collagen family. Cold Spring Harb Perspect Biol 3:a004978

Rich A, Crick FH (1961) The molecular structure of collagen. J Mol Biol 3:483–506

Rosenbloom J, Harsch M, Jimenez S (1973) Hydroxyproline content determines the denaturation temperature of chick tendon collagen. Arch Biochem Biophys 158:478–484

Ruggiero F, Exposito JY, Bournat P, Gruber V, Perret S, Comte J, Olagnier B, Garrone R, Theisen M (2000) Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. FEBS Lett 469:132–136

Ruggiero F, Koch M (2008) Making recombinant extracellular matrix proteins. Methods 45:75–85

Rutschmann C, Baumann S, Cabalzar J, Luther KB, Hennet T (2014) Recombinant expression of hydroxylated human collagen in *Escherichia coli*. Appl Microbiol Biotechnol 98:4445–4455

Seo N, Russell BH, Rivera JJ, Liang X, Xu X, Afshar-Kharghan V, Höök M (2010) An engineered α1 integrin-binding collagenous sequence. J Biol Chem 285:31046–31054

Shoulders MD, Raines RT (2009) Collagen structure and stability. Annu Rev Biochem 78:929–958

Smethurst PA, Onley DJ, Jarvis GE, O'Connor MN, Knight CG, Herr AB, Ouwehand WH, Farndale RW (2007) Structural basis for the platelet-collagen interaction: the smallest motif within collagen that recognizes and activates platelet Glycoprotein VI contains two glycine-proline-hydroxyproline triplets. J Biol Chem 282:1296–1304

Snellman A, Keränen MR, Hägg PO, Lamberg A, Hiltunen JK, Kivirikko KI, Pihlajaniemi T (2000) Type XIII collagen forms homotrimers with three triple helical collagenous domains and its association into disulfide-bonded trimers is enhanced by prolyl 4-hydroxylase. J Biol Chem 275:8936–8944

Song E, Yeon Kim S, Chun T, Byun HJ, Lee YM (2006) Collagen scaffolds derived from a marine source and their biocompatibility. Biomaterials 27:2951–2961

Stein H, Wilensky M, Tsafrir Y, Rosenthal M, Amir R, Avraham T, Ofir K, Dgany O, Yayon A, Shoseyov O (2009) Production of bioactive, post-translationally modified, heterotrimeric, human recombinant type-I collagen in transgenic tobacco. Biomacromolecules 10:2640–2645

Steplewski A, Majsterek I, McAdams E, Rucker E, Brittingham RJ, Ito H, Hirai K, Adachi E, Jimenez SA, Fertala A (2004) Thermostability gradient in the collagen triple helix reveals its multi-domain structure. J Mol Biol 338:989–998

Steplewski A, Hintze V, Fertala A (2007) Molecular basis of organization of collagen fibrils. J Struct Biol 157:297–307

Sutherland TD, Peng YY, Trueman HE, Weisman S, Okada S, Walker AA, Sriskantha A, White JF, Huson MG, Werkmeister JA, Glattauer V, Stoichevska V, Mudie ST, Haritos VS, Ramshaw JAM (2013) A new class of animal collagen masquerading as an insect silk. Sci Report 3:2864

Swatschek D, Schatton W, Kellermann J, Müller WE, Kreuter J (2002) Marine sponge collagen: isolation, characterization and effects on the skin parameters surface-pH, moisture and sebum. Eur J Pharm Biopharm 53:107–113

Sweeney SM, Orgel JP, Fertala A, McAuliffe JD, Turner KR, Di Lullo GA, Chen S, Antipova O, Perumal S, Ala-Kokko L, Forlino A, Cabral WA, Barnes AM, Marini JC, San Antonio JD (2008) Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. J Biol Chem 283:21187–21197

Tang B, Chiang TM, Brand DD, Gumanovskaya ML, Stuart JM, Kang AH, Myers LK (1999) Molecular definition and characterization of recombinant bovine CB8 and CB10: immunogenicity and arthritogenicity. Clin Immunol 92:256–264

Tang Y, Yang X, Hang B, Li J, Huang L, Huang F, Xu Z (2016) Efficient production of hydroxyl-ated human-like collagen via the co-expression of three key genes in *Escherichia coli* Origami (DE3). Appl Biochem Biotechnol [Epub ahead of print]

Toman PD, Pieper F, Sakai N, Karatzas C, Platenburg E, de Wit I, Samuel C, Dekker A, Daniels GA, Berg RA, Platenburg GJ (1999) Production of recombinant human type I procollagen homotrimer in the mammary gland of transgenic mice. Transgenic Res 8:415–427

Toman PD, Chisholm G, McMullin H, Giere LM, Olsen DR, Kovach RJ, Leigh SD, Fong BE, Chang R, Daniels GA, Berg RA, Hitzeman RA (2000) Production of recombinant human type I procollagen trimers using a four-gene expression system in the yeast *Saccharomyces cerevisiae*. J Biol Chem 275:23303–23309

Tomita M, Ohkura N, Ito M, Kato T, Royce PM, Kitajima T (1995) Biosynthesis of recombinant human pro-α1(III) chains in a baculovirus expression system: production of disulphide-bonded and non-disulphide-bonded species containing full-length triple helices. Biochem J 312:847–853

Tomita M, Kitajima T, Yoshizato K (1997) Formation of recombinant human procollagen I hetero-trimers in a baculovirus expression system. J Biochem 121:1061–1069

Tomita M, Munetsuna H, Sato T, Adachi T, Hino R, Hayashi M, Shimizu K, Nakamura N, Tamura T, Yoshizato K (2003) Transgenic silkworms produce recombinant human type III procollagen in cocoons. Nat Biotechnol 21:52–56

Vandersmissen L, De Buck E, Saels V, Coil DA, Anné J (2010) A *Legionella pneumophila* collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells. FEMS Microbiol Lett 306:168–176

Werkmeister JA, Ramshaw JAM (2012) Recombinant protein scaffolds for tissue engineering. Biomed Mater 7:012002

Williams KE, Olsen DR (2009) Matrix metalloproteinase-1 cleavage site recognition and binding in full-length human type III collagen. Matrix Biol 28:373–379

Xu Y, Keene DR, Bujnicki JM, Höök M, Lukomski S (2002) Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. J Biol Chem 277:27312–27318

Xu C, Yu Z, Inouye M, Brodsky B, Mirochnitchenko O (2010) Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability. Biomacromolecules 11:348–356

Xu X, Gan Q, Clough RC, Pappu KM, Howard JA, Baez JA, Wang K (2011) Hydroxylation of recombinant human collagen type I α1 in transgenic maize co-expressed with a recombinant human prolyl 4-hydroxylase. BMC Biotechnol 11:69

Yamada K, Yamaura J, Katoh M, Hata K, Okuda K, Yoshie H (2006) Fabrication of cultured oral gingiva by tissue engineering techniques without materials of animal origin. J Periodontol 77:672–677

Yang C, Hillas PJ, Báez JA, Nokelainen M, Balan J, Tang J, Spiro R, Polarek J (2004) The application of recombinant human collagen in tissue engineering. BioDrugs 18:103–119

Yasothornsrikul S, Davis WJ, Cramer G, Kimbrell DA, Dearolf CR (1997) viking: identification and characterization of a second type IV collagen in *Drosophila*. Gene 198:17–25

Yigit S, Yu H, An B, Hamaia S, Farndale RW, Kaplan DL, Lin YS, Brodsky B (2016) Mapping the effect of gly mutations in collagen on α2β1 integrin binding. J Biol Chem 291:19196–19207

Yoshizumi A, Yu Z, Silva T, Thiagarajan G, Ramshaw JAM, Inouye M, Brodsky B (2009) Self-association of *Streptococcus pyogenes* collagen-like constructs into higher order structures. Protein Sci 18:1241–1251

Yoshizumi A, Fletcher JM, Yu Z, Persikov AV, Bartlett GJ, Boyle AL, Vincent TL, Woolfson DN, Brodsky B (2011) Designed coiled coils promote folding of a recombinant bacterial collagen. J Biol Chem 286:17512–17520

Yu Z, Brodsky B, Inouye M (2011) Dissecting a bacterial collagen domain from *Streptococcus pyogenes*: sequence and length-dependent variations in triple helix stability and folding. J Biol Chem 286:18960–18968

Yu Z, Visse R, Inouye M, Nagase H, Brodsky B (2012) Defining requirements for collagenase cleavage in collagen type III using a bacterial collagen system. J Biol Chem 287:22988–22997

Yu Z, An B, Ramshaw JAM, Brodsky B (2014) Bacterial collagen-like proteins that form triple-helical structures. J Struct Biol 186:451–461

Zhang C, Baez J, Pappu KM, Glatz CE (2009) Purification and characterization of a transgenic corn grain-derived recombinant collagen type I α1. Biotechnol Prog 25:1660–1668

Zhang H, Fan D, Deng J, Zhu C, Hui J, Ma X (2014) Effect of Tris-acetate buffer on endotoxin removal from human-like collagen used biomaterials. Mater Sci Eng C Mater Biol Appl 42:124–129