Subcellular Biochemistry 82

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

# Fibrous Proteins: Structures and Mechanisms

Springer

# Subcellular Biochemistry

Volume 82

**Series editor**
J. Robin Harris
University of Mainz, Mainz, Germany

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



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

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

ISSN 0306-0225
Subcellular Biochemistry
ISBN 978-3-319-49672-6 ISBN 978-3-319-49674-0 (eBook)
DOI 10.1007/978-3-319-49674-0

Library of Congress Control Number: 2016963615



Printed on acid-free paper

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

# 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 newcomers 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



D.A.D. Parry, J.M. Squire (eds.), *Fibrous Proteins: Structures and Mechanisms*, Subcellular Biochemistry 82, DOI 10.1007/978-3-319-49674-0_1

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, polynucleotides, 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 polypeptides, 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 $NH_2$-$C_\alpha$HR-C′OOH. The core of the amino acid is the alpha carbon ($C_\alpha$), which is a tetrahedral carbon atom making single bond interactions 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 characteristic 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 differences 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, connective 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

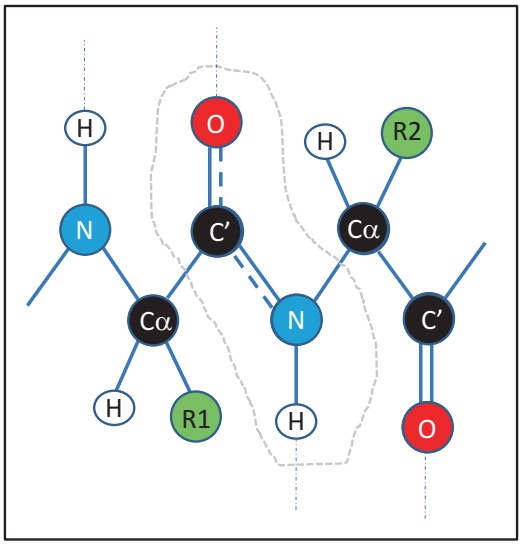


**Fig. 1.1** Schematic diagram of two amino acids ($NH_2.C_\alpha HR.C'OOH$) linked by a peptide bond. The C′OOH group of one amino acid and the $NH_2$ group of its successor bond through a condensation reaction where the hydroxyl group from the C′ atom and a hydrogen from the amine group are eliminated 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 $C_\alpha$, 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

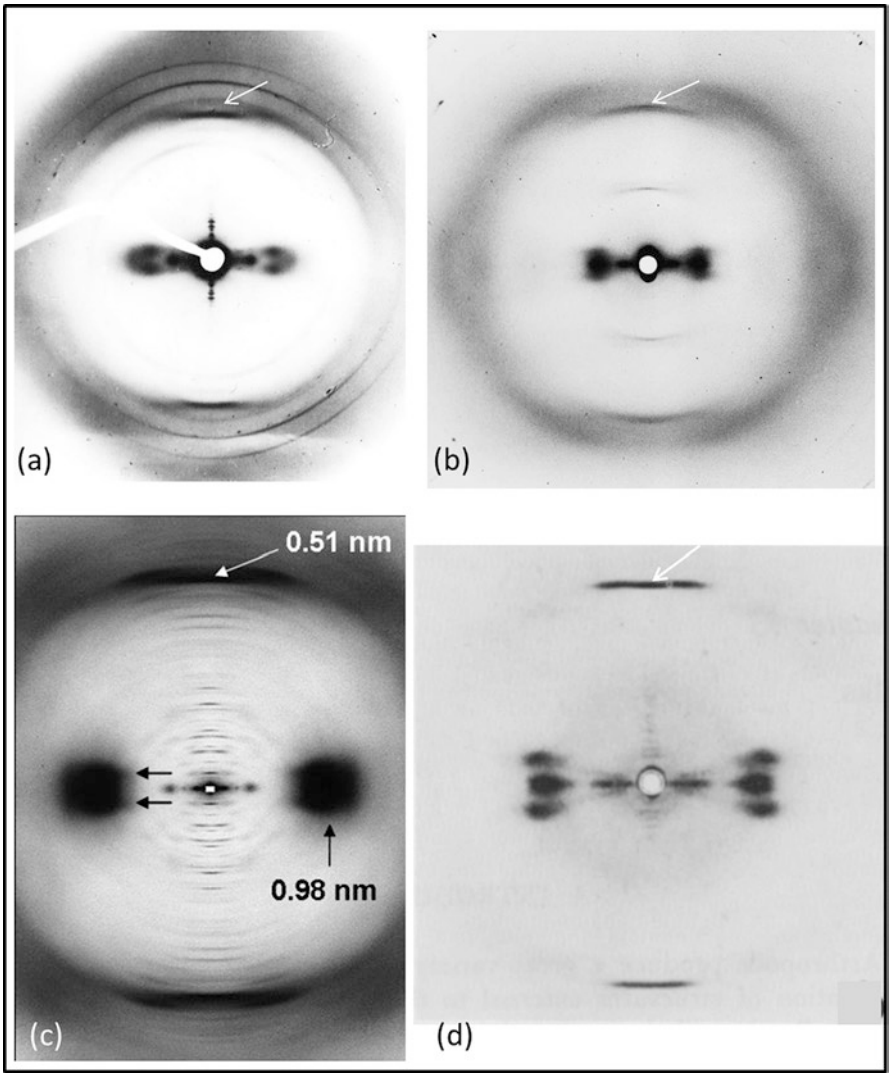


**Fig. 1.3** High-angle X-ray diffraction patterns from a number of fibres with the α–helical conformation: (**a**) paramyosin filaments in molluscan muscle, (**b**) poly-γ-benzyl-L-glutamate, (**c**) α–keratin (from Parry et al. 2008; taken by T.P. MacRae) and (**d**) α–silk from the honeybee (*Apis 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,

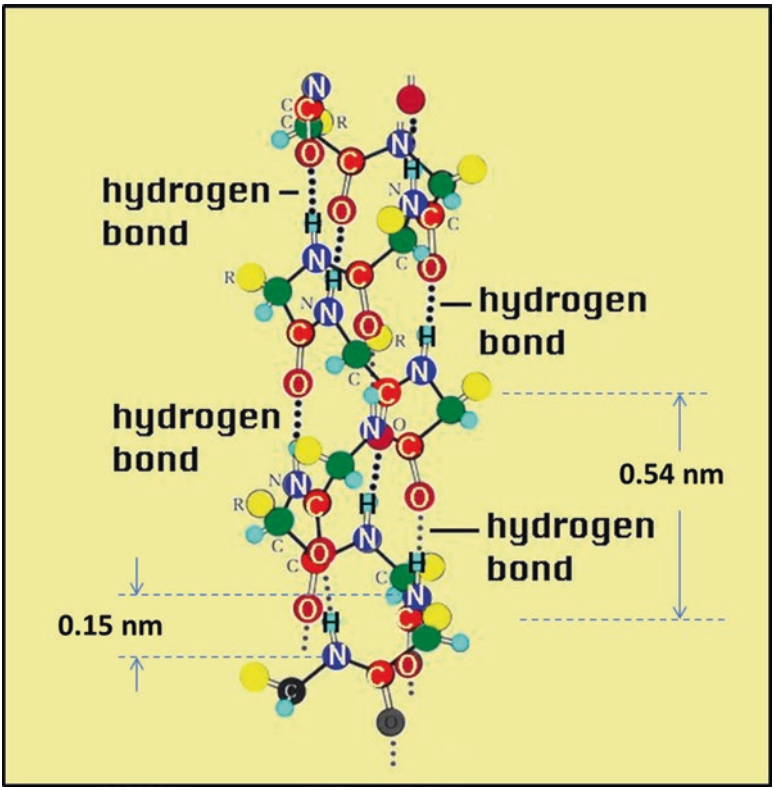


**Fig. 1.4** Diagram of the α–helix structure showing the 0.54 nm helix pitch, the 0.15 nm axial translation between successive amino acids and the hydrogen-bonds that lie almost parallel to the helix axis (Adapted from http://academic.brooklyn.cuny.edu/biology/bio4fv/page/alpha_h.htm)

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

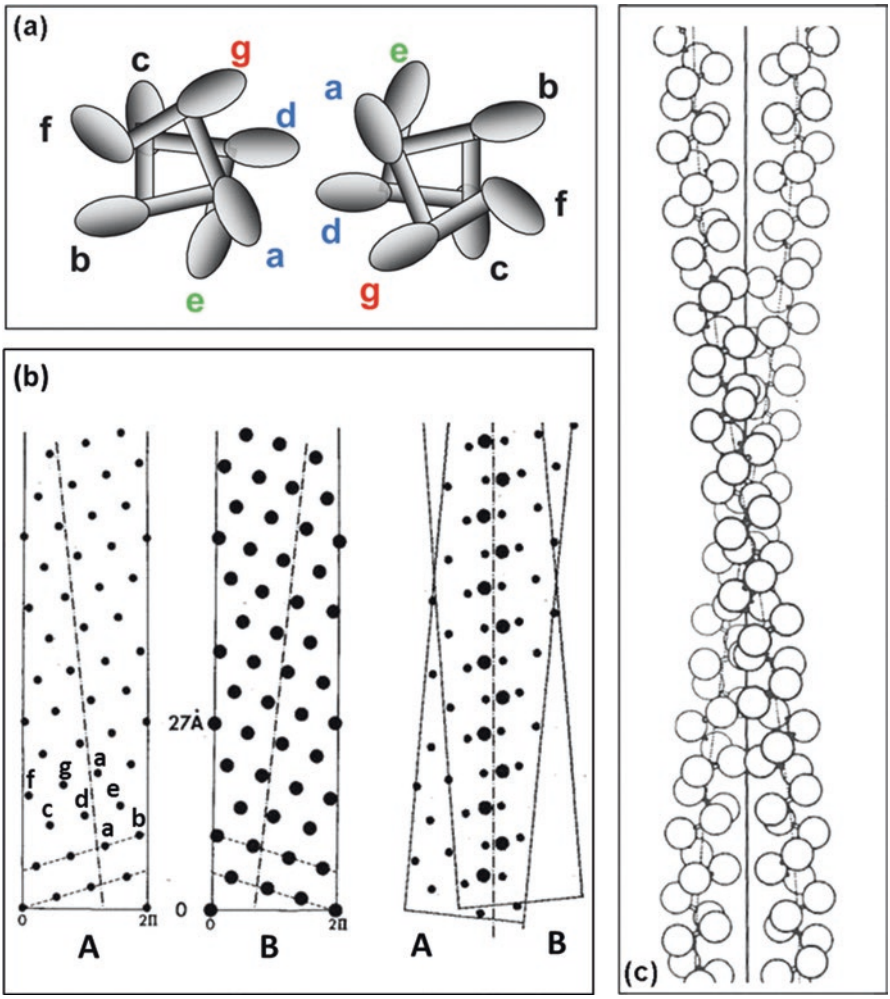


**Fig. 1.5** (**a**) Schematic diagram showing the **a** to **g** heptad positions in two adjacent two-chain coiled coils. Apolar residues usually occur in positions **a** and **d**. Charged groups in positions **e** and **g** can determine the polarity of the two chains and their relative axial stagger. 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).

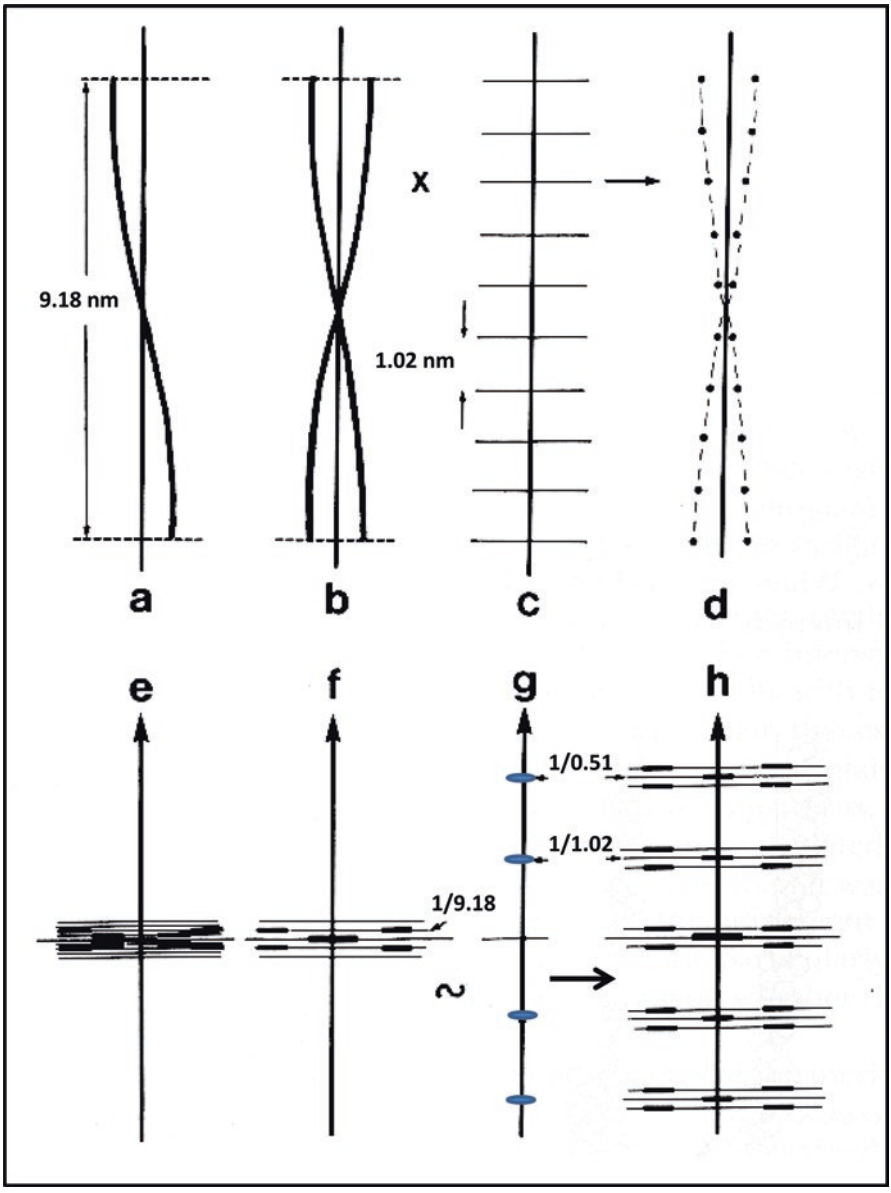


**Fig. 1.6** Generation of the form of the diffraction pattern from a two-chain coiled-coil α–helical molecule. The pitch of the coiled coil is taken here to be 18.36 nm and the subunit repeat along each strand is 18.36/18 = 1.02 nm. The diffraction pattern (**e**) of a single continuous helical strand (**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 $m/1.02\ \text{nm}^{-1}$ ($m$ 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))

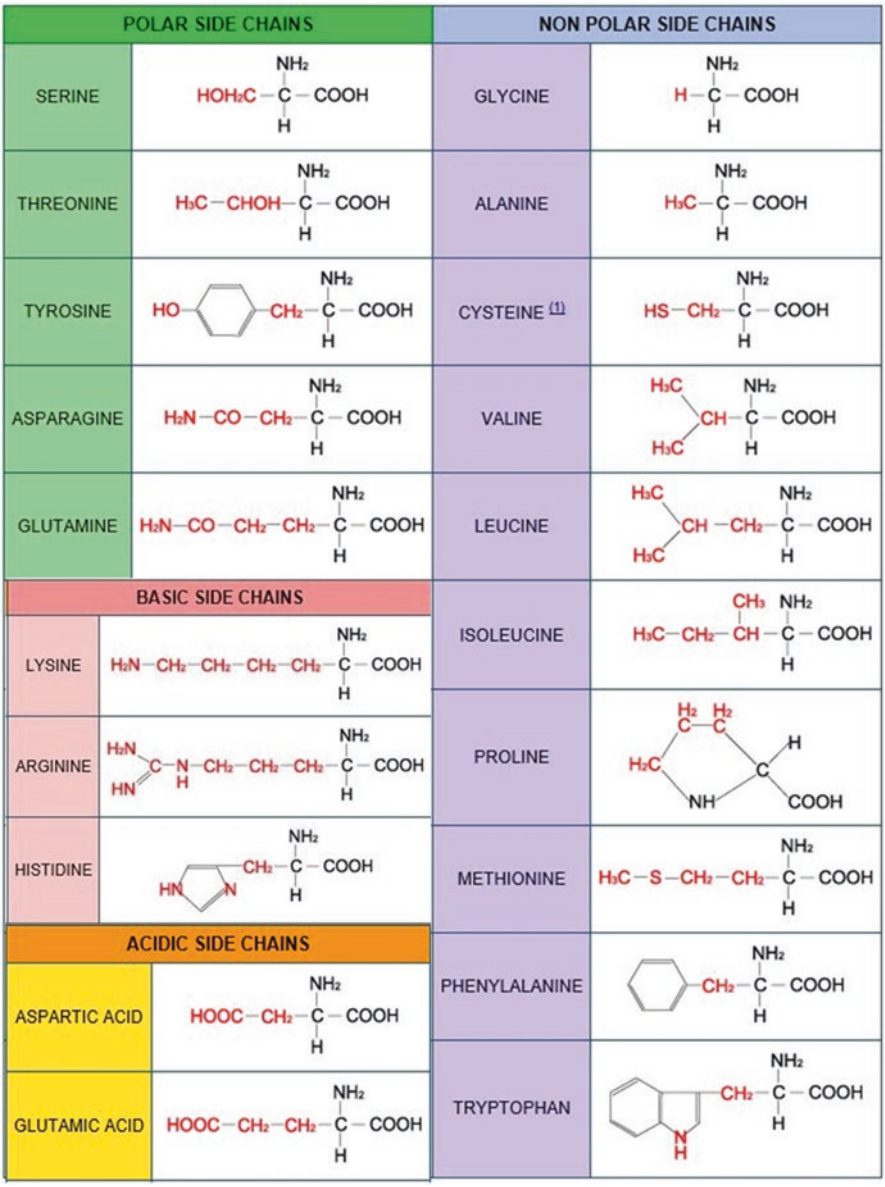


**Fig. 1.7** The 20 amino acids that commonly occur in proteins are grouped according to their chemical properties (For details see text. Adapted from: http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/formuleAA/)

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

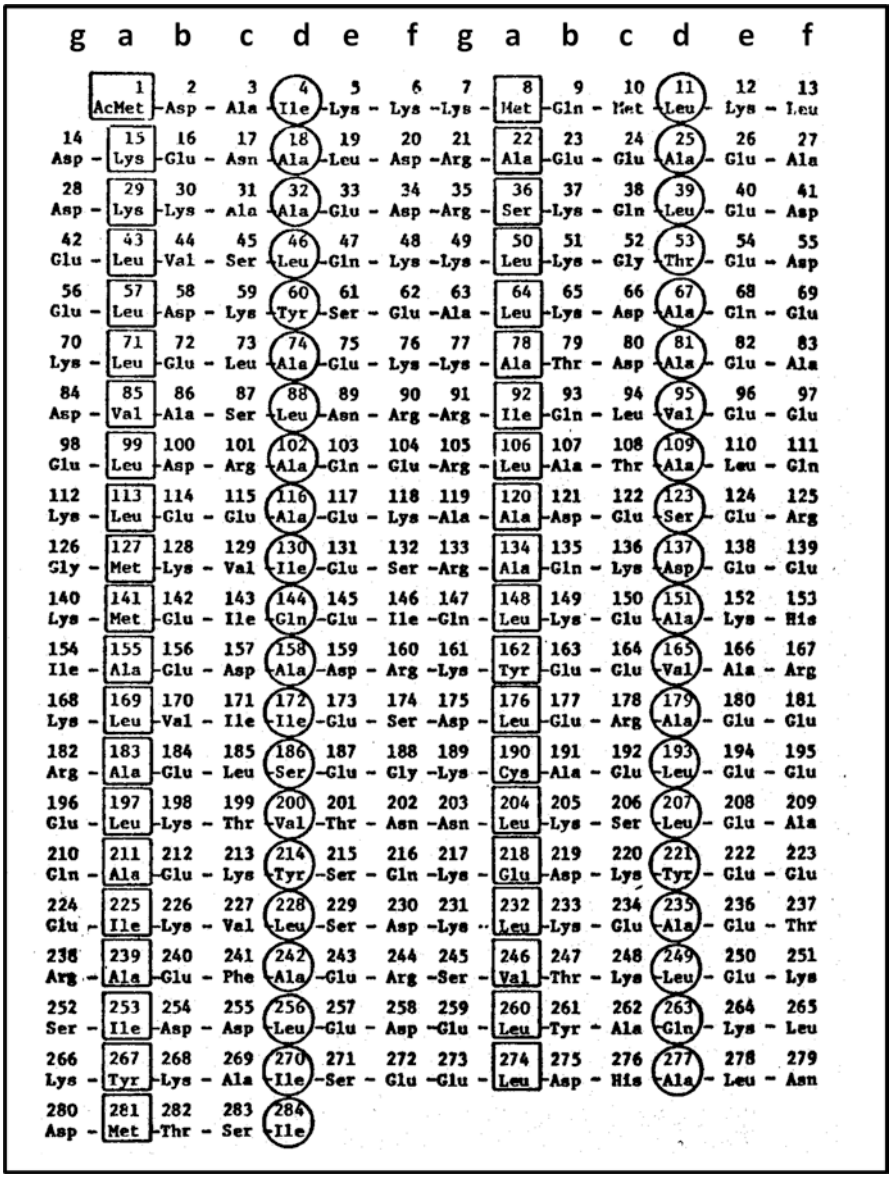


**Fig. 1.8** The amino acid sequence of α–tropomyosin from rabbit skeletal muscle according to Stone et al. (1975). The residues have been grouped as two sets of seven and labelled with the letters **a** to **g** to show the heptad structure in which hydrophobic residues tend to occupy the **a** and **d** positions. The three letter amino acid codes are given in Fig. 1.9. (Reproduced from Squire (1981; page 186))

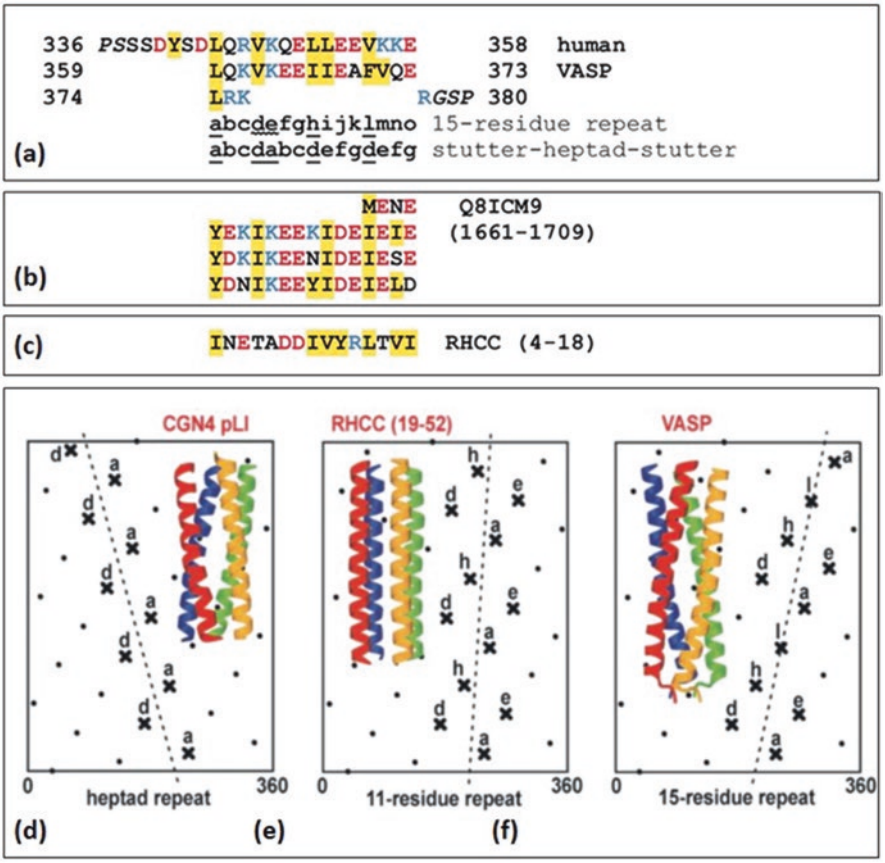


**Fig. 1.9** (**a**) The amino acid sequence of the vasodilator-stimulated phosphoprotein (VASP) tetramerization domain (TD) has been formatted to highlight the 15-residue repeat pattern. Apolar residues and Y and F residues are highlighted in *yellow*, acidic residues in *red* and basic residues in *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

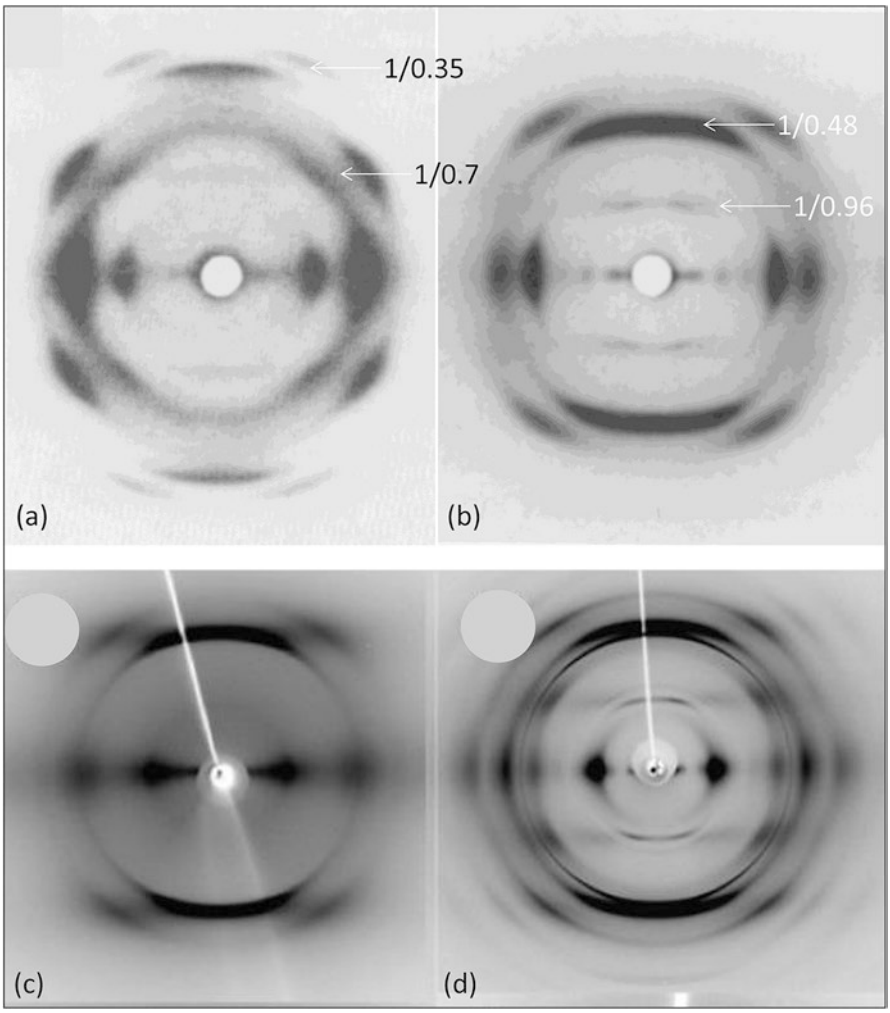


**Fig. 1.10** Typical X-ray diffraction patterns from β-structures: (**a**) β-silk, (**b**) cross-β structure, (**c**, **d**) amyloid proteins. (**c**, **d**) (From Serpell (2013))

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.

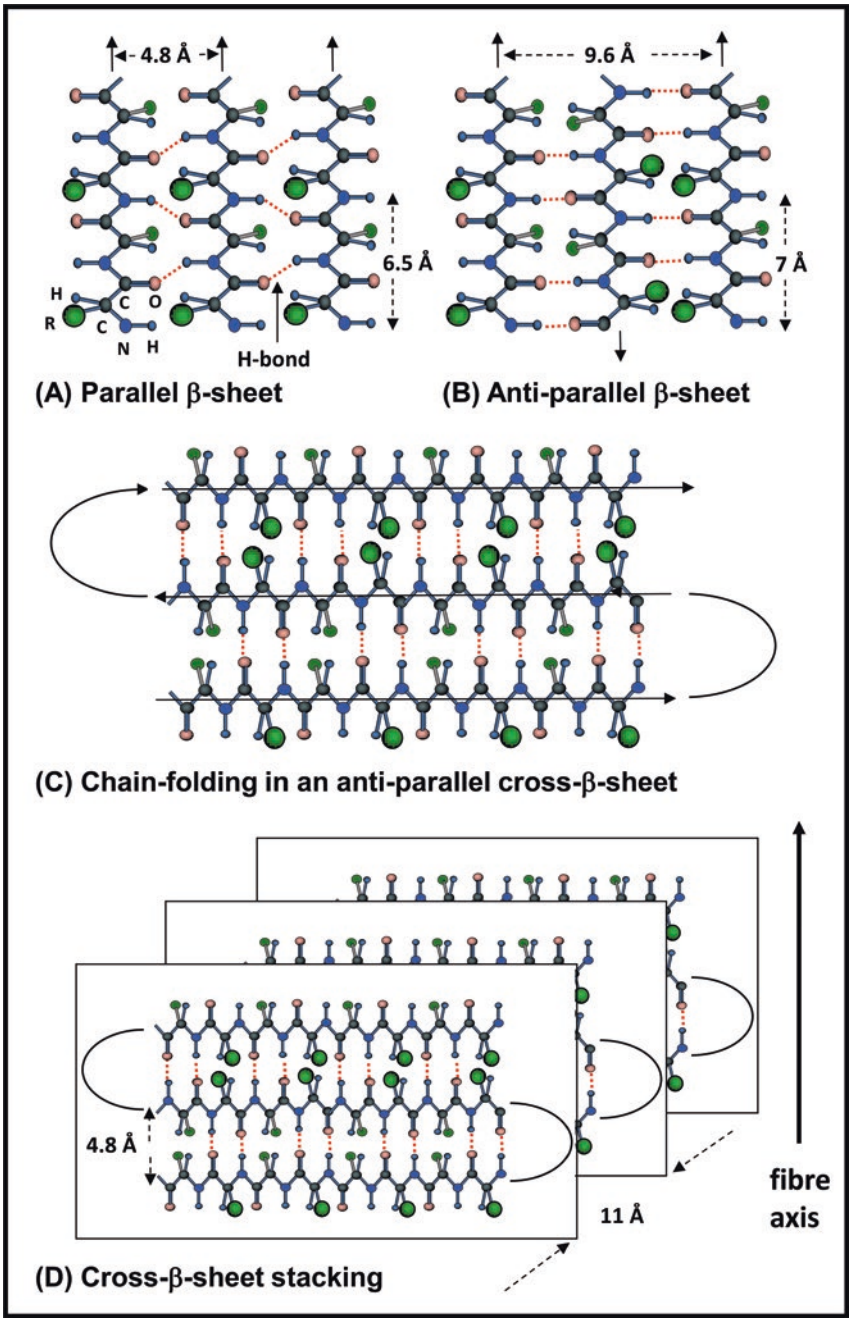


**Fig. 1.11** The basic arrangements of β-strands in hydrogen-bonded β-sheets; (**a**) parallel chains, (**b**) antiparallel chains. *Green spheres* of different sizes denote sidechain groups directed either towards (*large spheres*) or away from (*small spheres*) from the reader. Hydrogen-bonds are 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

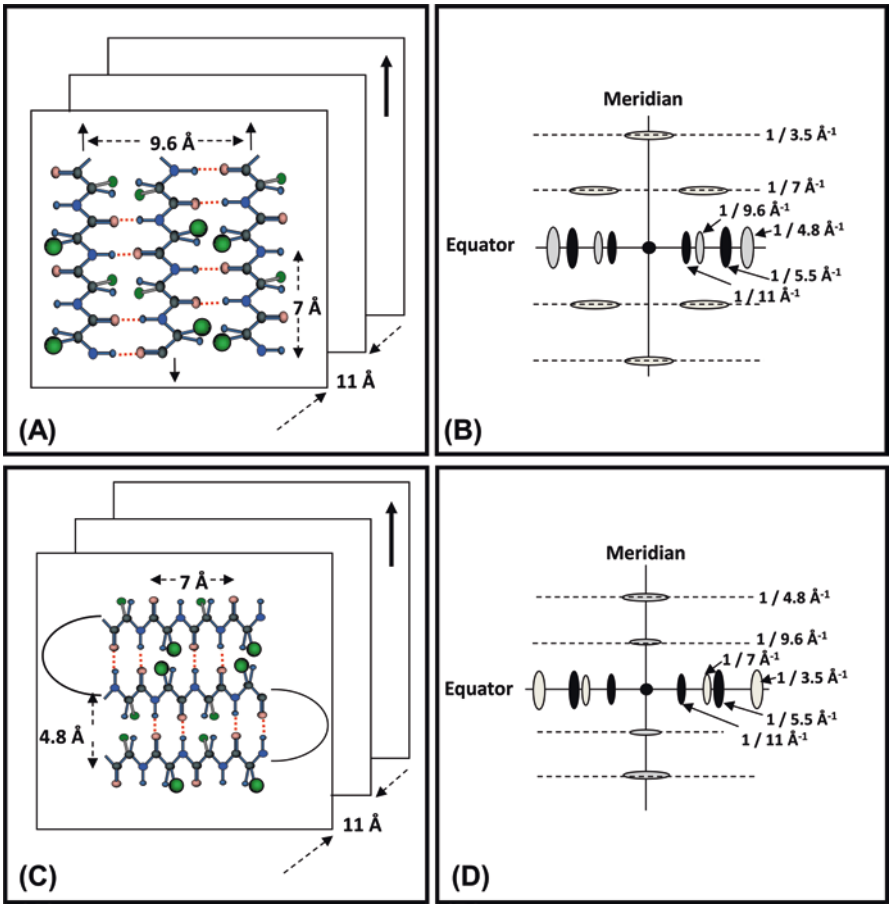


**Fig. 1.12** The differences that might be observed in the fibre diffraction patterns from oriented samples of antiparallel β-structures depending on whether the chains are aligned along (**a**, **b**) or perpendicular to (**c**, **d**) the fibre axis. Colour coding on (**a**, **b**) as in Fig. 1.11. 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 $(\text{Gly-Pro-Hyp})_n$ 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).