Subcellular Biochemistry 93

J. Robin Harris
Jon Marles-Wright *Editors*

Macromolecular Protein Complexes II: Structure and Function

Springer

Subcellular Biochemistry

Volume 93

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

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

J. Robin Harris · Jon Marles-Wright
Editors

Macromolecular Protein Complexes II: Structure and Function

Springer

Editors
J. Robin Harris
Institute of Zoology
University of Mainz
Mainz, Germany

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

ISSN 0306-0225 ISSN 2542-8810 (electronic)
Subcellular Biochemistry
ISBN 978-3-030-28150-2 ISBN 978-3-030-28151-9 (eBook)
https://doi.org/10.1007/978-3-030-28151-9

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

Preface

The compilation of this volume of the *Subcellular Biochemistry* series follows on from several volumes already devoted to the Macromolecular Protein Complex theme. The present, and indeed subsequent, volumes are justified because so many interesting relevant topics, deriving from X-ray crystallographic, cryo-EM and other structural studies, have appeared in the scientific literature over recent years. While already building up to be an almost encyclopaedic coverage, there is still much that could be added.

This present volume of the series contains a diverse collection of 19 fascinating chapters, as can be seen from the content list. While we will not expand here on any one chapter, it is clear that all the chapters stand alone, with some chapters complementing each other. The overall field of Macromolecular Protein Complexes continues to expand due to developments in cryogenic transmission electron microscopy and computational methods for reconstructing single particles and volumes. It therefore represents an important continuing component of structural biology, without in anyway detracting from the importance of X-ray crystallographic and NMR studies on an increasingly large number of smaller, usually monomeric, protein molecules.

The receipt and compilation of the chapters in this book have been a true excitement. Chapter authors have provided much exciting material for inclusion in the book, many containing impressive figures. As so often happens when editing a multi-author book, several planned chapters were either withdrawn or failed to appear in time for inclusion. Nevertheless, ample material is available to create a strong book, to supplement Volumes 82, 83, 84, 87 and 88 of the series (see https://www.springer.com/series/6515?detailsPage=titles). With two further related volumes at an earlier stage of compilation, continuation has been established, and it is hoped that many more interesting related topics will eventually be included in the *Subcellular Biochemistry* series.

The book is directed towards advanced undergraduates, postgraduates, research workers and academics who have a specialism or broad interest in molecular structural biology. It is hoped that the book, available in print and as an e-book, will be of interest and value to many.

Mainz, Germany — J. Robin Harris
Newcastle upon Tyne, UK — Jon Marles-Wright
June 2019

Contents

Chapter 1
Introduction: Protein Oligomerization and the Formation of Macromolecular Assemblies

J. Robin Harris and Jon Marles-Wright

Abstract The ability of biomolecules to link together to form higher order assemblies underlies much of cellular structure and function. Here we emphasise protein oligomerisation and discuss some of the principles of molecular interaction, from early considerations through to the present day. A few protein examples are presented, selected from our research interests, to highlight assembly features, ranging from the hemoglobins, the hemocyanins to the peroxiredoxins, collagen, the encapsulins and ferritins.

Keywords Protein · Oligomerization · Assembly · Hemoglobin · Hemocyanin · Peroxiredoxin · Collagen · Encapsulin · Ferritin

Fundamentals

A fundamental property of the organic molecules present in biological organisms is their ability to group together to form polymer chains, as found in DNA, proteins and carbohydrates, or bilayer sheets as for the lipids. Furthermore, these four principal biomolecules also interact with one another. The ability of biological molecules to link or associate together is implicit in the structure of DNA, with anti-parallel helices containing covalently linked nucleotide phosphates, dimerized as a *double helix* stabilized by hydrogen bonds between the complementary A-T, G-C base pairs. Indeed be it DNA or a protein, the dimer can be looked upon as the simplest oligomeric form; in the case of the helix it does not expand beyond the dimer (except under crystallization conditions, and in Holliday junctions and

J. R. Harris (✉)
Institute of Molecular Physiology, University of Mainz, Joachim-Becher-Weg 7, 55099 Mainz, Germany
e-mail: rharris@uni-mainz.de

J. Marles-Wright
School of Biology, Newcastle University, Newcastle upon Tyne NE2 4HH, UK
e-mail: Jon.marles-wright1@ncl.ac.uk

J. R. Harris and J. Marles-Wright (eds.), *Macromolecular Protein Complexes II: Structure and Function*, Subcellular Biochemistry 93,
https://doi.org/10.1007/978-3-030-28151-9_1

G-quadruplexes). For protein molecules the situation is somewhat different, with a multitude of structural possibilities for the creation of higher oligomeric forms.

Starting with the sequence of the 22 proteinogenic amino acids in a polypeptide chain, it is immediately apparent that the number of variations of primary structure is enormous, in turn reflected in the secondary structure with its combination of alpha helices, beta sheets, and connecting loops, that then determine the folding of the polypeptide chain and tertiary structure, stabilized by electrostatic/ionic, hydrogen bonds and in some cases by disulphide bonds. The arrangement of protein chains into multimeric assemblies gives the quaternary structure.

There are cases where secondary structure elements from individual chains in a multimeric-assembly combine to form tertiary structures across subunits. For instance, in some fibrous proteins, the tertiary structure includes the coming together of a number of closely associated polypeptide chains in the form of double or triple coiled-coil alpha helices, as in the collagen heterotrimer. The coming together of beta sheets from a number beta barrels has also emerged as a significant structural feature. This is especially apparent in the amyloid proteins, with their extensive cross-beta sheet and beta-helix structures formed by Tau filaments (Fitzpatrick et al. 2017) and amyloid fibres of beta-macroglobulin (Iadanza et al. 2018).

The definition of individual *domains* within a protein also has significance in relation to dimerization and multimerization. To some extent these features immediately indicate the possibility of creating sequence-defined protein-protein "subunit" interactions, that can then determine the widely occurring quaternary structure of multimeric macromolecules. Within protein multimers, the overall array of inter-subunit contacts and interactions is always determined by accessible ionic and hydrophobic surface groups rather than those of internally *buried* amino acids; nevertheless these latter determine the internal and overall surface geometry of the subunit. Whatever the protein-protein interaction involved in determining tertiary and quaternary structure, it generally does not involve covalent bonding (unlike in primary structure, and the relatively labile disulphide bonds in tertiary structure). It hardly needs to be emphasised that this leads to an increasing number of structural possibilities, exemplified by the expanding number of protein structures that have been determined by X-ray crystallography over the past 50 years (www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=tota).

Among these are numerous large multimeric (homo- and hetero-oligomeric) protein complexes, many of which were *discovered* or defined by transmission electron microscopy, that have until relatively recently presented difficulties for crystallization and X-ray crystallography. The coming together of conventional transmission electron microscopy and cryo-electron microscopy with X-ray crystallography has led to much progress in the structural analysis of many large oligomeric proteins and macromolecular assemblies. Nevertheless, these structural tools are greatly supported by several essential biochemical and biophysical techniques, such as electrophoresis, chromatography, centrifugation, light scattering, spectroscopy and others, for the purification and analysis of proteins (see three recent texts "Protein Purification and Analysis", iConcept Press). The chapters in

the present book and related volumes of the *Subcellular Biochemistry* series present detailed reviews on the structure and function of many interesting protein complexes, and thereby create an almost encyclopaedic coverage of this important aspect of protein biochemistry.

In a class of their own are the non-covalent antigen-antibody complexes, so important for the immune response. Furthermore, the polyclonality of in vivo antibody production contrasts with the monoclonality of in vitro cellular production, along with the varying affinity of antibodies for their highly specific protein epitope. Whilst being an exciting area of study, this is beyond the scope of the present chapter and this particular volume. Likewise, a consideration of covalent protein-protein and protein-nucleic acid complexes, whilst of great interest, cannot be dealt with in this chapter.

Here we present a limited selection of macromolecules in which we are interested and have studied in recent years, purely as examples of the many macromolecular assemblies that could be considered. Undoubtedly each protein complex possesses unique structural features responsible for the quaternary structure created by the subunit protein-protein association. Nevertheless, some principles of assembly can be defined and have been advanced by protein chemists, virologists and others. In the main, however, symmetrical and asymmetrical protein complexes remain as individual items for study, with an ever increasing number under investigation as present-day biological and biomedical research advances.

Protein Oligomeric Structure: Background

Ahnert et al. (2015) advanced a classification model for protein complexes based upon theoretically possible quaternary structure organization. Although a *periodic table* of protein subunit organization was established as a useful predictive model, it serves primarily to emphasise the incredible number of quaternary structures that spontaneously occur in nature. There can be no doubt that protein complex formation occurred during early evolution (Marsh and Teichmann 2014) and that the accrued functional benefits resulting from stability, allostery, and the conservation and compartmentalisation of reaction intermediates, have been retained, carried forward and *evolved* within all species through to the present day.

The late N. Michael Green, who worked at the UK MRC National Institute for Medical Research, Mill Hill (http://www.bridgesmathart.org/art-exhibits/bridges06/green.html), illustrated the diversity of protein multimerization through "plastic monkey" models, which made 'asymmetric units' in his sculptures. His wide interest in protein structure, including immunoglobulins, avidin, glutamine synthase, and transcarbamylase, informed his ideas and art (Valentine and Green 1967; Green et al. 1972; Green 1972, 1990).

"*Every living organism produces a variety of complex structures by self-assembly of identical building blocks of one or more types. These structural units are usually protein molecules which have evolved to assemble spontaneously, using multiple weak bonds. The principles of thermodynamics ensure that the most stable links are used. The number of such links is maximised in symmetrical structures, in*

which all the units are identically bonded. Natural selection ensures that assemblies with useful biological properties are perpetuated (e.g. *virus coats, muscle fibres, mitotic spindles, intercellular junctions and many enzymes*)."

Whilst being an over simplification of the complex inter-molecular bonding present in any symmetrical oligomer, along with asymmetrical oligomers, Green's early concepts did indicate the basic principle of homo-oligomeric and hetero-oligomeric assembly, insight that has been advanced greatly in recent years. Further examples of early application of negative staining to the enzymes complexes and macromolecules was presented by Van Bruggen et al. (1960, 1981), Matthews and Bernhard (1973), and Oliver (1973).

The Structural Techniques

Biochemical and biophysical techniques have contributed toward the study of protein structure. Whilst analytical ultracentrifugation played a useful role in defining monomers and their higher mass assemblies during the first half of the 20th century, its contribution has diminished. However, first and foremost must be native- and SDS-polyacrylamide gel electrophoresis (PAGE), techniques valuable for the assessment of protein purity and subunit composition (Laemelli 1970). These electrophoretic techniques go hand in hand with the various chromatographic, preparative centrifugation and spectroscopic procedures used to achieve and assess protein purity from cells and biological tissues, or more usually in recent years from proteins expressed in bacteria, yeasts and animal cells. In this context, gene sequencing and cloning, along with protein amino acid sequencing, have played an important role. Indeed, the exponential increase in protein structure depositions in the protein databank has been a direct consequence of the technological advances from the major structural genomics initiatives over the last twenty years (Grabowski et al. 2016) and technological advances in X-ray sources, both in the laboratory and at synchrotron facilities (Owen et al. 2016). Laboratory automation and robotics have been of particular impact, with nano-scale protein crystallisation robotics reducing the amount of precious protein required for crystallisation, and automated sample-changers at synchrotrons speeding up data collection from protein crystals.

For the study of small proteins (<250 kDa), in addition to protein sequencing, X-ray crystallography (and to a lesser extent NMR spectroscopy) (Bernal and Crowfoot 1934) has been the predominant technique, expanding seemingly exponentially ever since the atomic structures of myoglobin and haemoglobin were determined more than 50 years ago (Kendrew et al. 1960; Perutz et al. 1960). With the polypeptide hormone insulin, the presence of two polypeptide chains, containing three covalent disulphide bridges was defined by Fred Sanger and colleagues (Ryle et al. 1955). Insulin Zinc-crystals contain hexamers as the unit cell (Harding et al. 1966), indicating the property of molecules to non-covalently associate under appropriate solution conditions. Indeed, the repeating unit cell

within a protein crystal is often an oligomer, thereby indicating the strong tendency of like molecules to associate. It is abundantly clear that the progressive increase in computing power through the recent decades has greatly helped protein crystallography for the analysis of protein structure, as it has in many areas of science and technology. However, for an increasing number of large protein molecules the production of suitably diffracting crystals and subsequent generation of crystallographic data has proved a stumbling block.

It is here that transmission electron microscopy has made a significant contribution. Starting with negative staining and to a lesser extent metal shadowing that reveal the protein surface profile (reviewed by Harris 2015), followed by higher resolution cryo-electron microscopy of unstained vitrified specimens (Adrian et al. 1984) (that also has an increasing potential to reveal protein internal and surface structure to the low Å atomic level), the structure of many protein oligomers and assemblies has been defined.

The low resolution achieved by negative staining (~20 Å) has been steadily improved with the increasing availability of cryo-electron microscopic data, derived from the technical advances at the instrumental level, particularly in the development of direct electron detectors (McMullan et al. 2016), and massively increased image processing capability of recent decades, to achieve near-to atomic resolution (~2 Å) from single particle analysis and crystallographic analysis of 2D membrane crystals (Henderson 2015; Vinothkumar and Henderson 2016; Stahlberg et al. 2015). The development of software for protein digital image analysis and 3D reconstruction has played a significant role in this (Zivanov et al. 2018), along with the fitting/docking of higher resolution X-ray data of subunits into lower resolution electron microscopy data and homology modelling (Nicholls et al. 2018). It is appropriate to acknowledge the contribution of numerous physical/biophysical and computer-literate scientists and electron microscopists to the development of this field; it probably would not have happened in the hands of biologist, biochemists and protein chemists alone! Indeed, the continuing impact of computer refinement, image analysis and 3D modelling on the understanding of protein assemblies is steadily increasing, but will not be dealt with in the present chapter, and it does not represent a major component of the book as a whole.

Some Oligomeric Proteins and Complexes

Hemoglobin

Following the seminal X-ray crystallographic studies on myoglobin and haemoglobin by Kendrew and Perutz protein biochemists must have immediately contemplated as to why myoglobin is a monomer and the evolutionary-related haemoglobin A is an α2ß2 globin tetramer? And, why is it not and α4 or ß4 tetramer? The expanding sequence data on the embryonic, foetal and adult

hemoglobins and the range of hemoglobinopathies has generated much information as to the range of tetramer formation, all indicating that this is the predominant and stable oligomeric form adopted by the haemoglobin monomers. The amino acid substitution of valine in the place of glutamine, at position 6 of the ß-globin chain due to the single codon missense mutation GAG to GTG, is responsible for sickle cell haemoglobin (HbS). This valine substitution creates a more hydrophobic environment and increases the tetramer instability when deoxygenated, resulting in decreased solubility of the HbS, with formation of helical polymers/fibres that bundle within the erythrocyte, essentially in a *quasi* crystalline manner. Due to the restricted space inside the erythrocyte, the HbS bundles bend, creating the characteristic sickle cells. Repeated cycles of oxygenation of HbS oxygenation and deoxygenation lead to irreversible sickling. This intracellular *quasi* crystallization of HbS is one of the few examples of natural 3D crystallization found in nature, other examples include the potato proteinase inhibitor (Rodis and Hoff 1984), the *Bacillus thuriengensis* toxin (Li et al. 1991) and the polyhedrin proteins from insect viruses (Ji et al. 2015). There are rather more examples of naturally occurring 2D membrane protein crystals (*Halobacterium halobum* purple membrane and the chloroplast thylakoid light harvesting complex).

The annelid, crustacean and insect hemoglobins all exhibit rather different oligomerization properties to those from vertebrates, exemplified most strongly by the earthworm (*Lumbricus tearrestris*) and ragworm (*Nereis virens*) hexagonally-ordered high molecular mass hemoglobins (erythrocruorins) (Fig. 1.1). Although earthworm haemoglobin was one of the first proteins from which 3D crystals were produced (and indeed it readily produces 2D crystals), only a low resolution 5.5 Å X-ray structure has been produced (Royer et al. 2000). and it has fallen to cryoEM of single molecules to generate singe particle reconstructions. Chen et al. (2015) produced an 8.1 Å structure and Afanasyev et al. (2017) achieved a higher resolution structure (at ~3.8 Å). In this hexameric double-layer macromolecule a significant feature is the 1/12th unit containing 12 haem-containing globins and a "stem" containing the triple coiled-coil linker chain (Fig. 1.2) directed to the centre of the hexagonal structure (PDBID: 5M3L). Other invertebrate haemoglobins are apparently less well-ordered oligomers (Rousselot et al. 2006) and even occur as linear haemoglobin polymers (Borhani et al. 2012).

The Haemocyanins

For the megadalton copper-containing cephalopod and gastropod respiratory proteins, the haemocyanins, ring-like decameric ~3.8 MDa oligomers are formed (Fig. 1.3), and in some species these form as didecamers and multidecamers (Markl 2013; Kato et al. 2017). Negative staining in the presence of PEG potentiates the formation of 2D hemocyanin crystals (see Fig. 1.3) (Harris and De Carlo 2014). An antiparallel subunit dimer (protamer) is the repeating unit within the decamer (Harris et al. 2004; Meissner et al. 2007a). The individual elongated haemocyanin

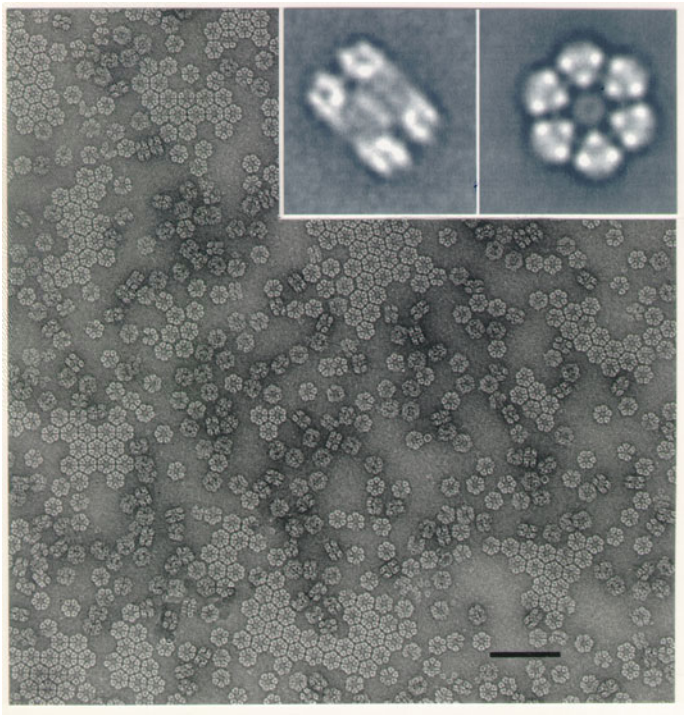

Fig. 1.1 Haemoglobin from the ragworm *Nereis virens*. The hexameric protein complex is shown by cryo-negative staining (courtesy of the late Marc Adrian), with the molecules oriented in the two predominant orientations. The insets show 2D molecular averages

subunits contain seven or eight covalently-linked genetically individual "functional units" (FUs), each with a copper-containing oxygen binding site. Gene duplication during evolution is believed to have generated these distinct FUs. Indeed, the polypeptide linker regions between the FUs can be considered to create an "oligomeric" subunit containing multiple covalent links. Most molluscan hemocyanins will dissociate to the single subunit in an alkaline buffer, but the hemocyanin subunit dimer from the chiton *Acanthochiton fasicularis* remains stable (Harris et al. 2004). An understanding of the formation of hemocyanin multidecamers of varying length (and indeed the proteolytically derived helical tubules), together with the deviant tridecameric mega-hemocyanin (Gatsogiannis et al. 2015) remain to be established.

Recent, 3 Å, crystallographic data from the *T. Pacificus* haemocyanin decamer (Gia et al. 2015) has revealed considerable structural detail. The subunit organization and copper-containing FUs were defined, in particular the hydrophobic FU interaction sites, responsible for creation of the protomer (dimer). Calcium/ionic bridging between dimers is also thought to be of importance for the creation of the higher order decamers.

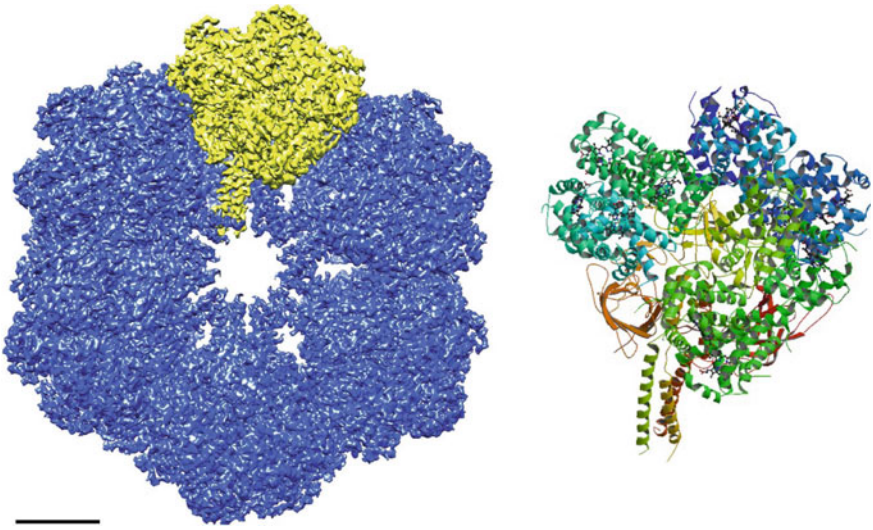

Fig. 1.2 The asymmetric unit ('protomer' or '1/12th unit') of the earthworm *Lumbricus terrestris* haemoglobin at 3.8 Å containing 12 haem-containing globins and a "stem" containing the triple coiled-coil linker chain (PDB entry 5m3l) from cryoEM (Afanasyev et al. 2017). Reproduced with permission of the International Union of Crystallography. (https://journals.iucr.org/)

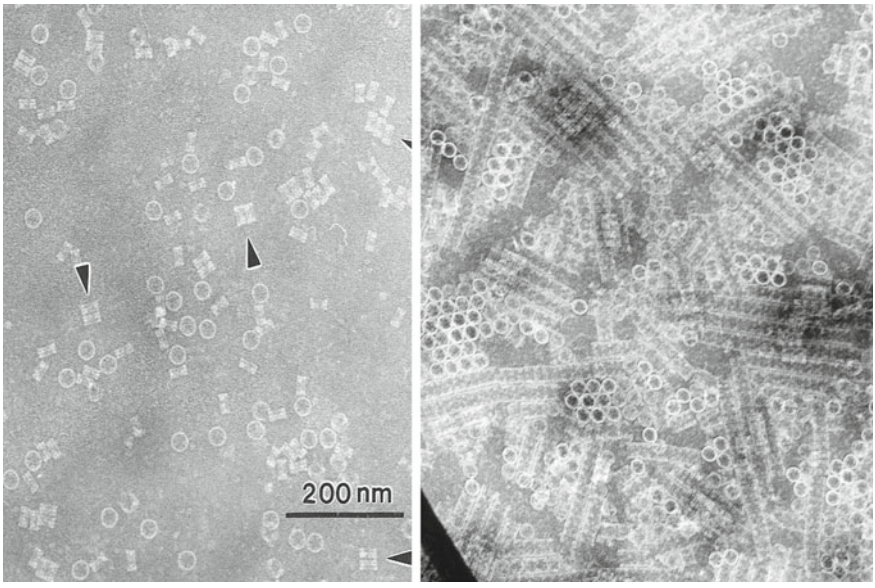

Fig. 1.3 TEM images of octopus haemocyanin decamers negatively stained with ammonium molybdate-trehalose (LHS) and ammonium molybdate-trehalose in the presence of PEG 1000 (RHS) across a hole. Note the ring-like face-on decamers and the side-on rectangular molecular images that have a tendency to form stacks (arrows). The presence of PEG during the negative staining procedure with a higher protein concentration promotes the formation of 2D crystals (De Carlo and Harris 2011)

The oligomeric haemocyanins from numerous other arthropods, products of a different gene family to the molluscan haemocyanins, are currently under active investigation and present an equally fascinating structural story, yet to be fully revealed (Martin et al. 2007).

The Peroxiredoxins

The large family of thiol-specific antioxidant and redox signalling enzymes, generally termed the peroxiredoxins, has emerged as a significant component of the overall enzymic antioxidant protection. The discovery of the erythrocyte peroxiredoxin-II (Prx-II_as a ring-like protein oligomer goes back to the late 1960s, when a ring-like decameric protein of then unknown function was defined by negative stain electron microscopy of protein extracts produced from erythrocyte ghosts (Harris 1968, 1969). Subsequently, from SDS-PAGE the single subunit molecular mass of ~20 kDa was defined (Harris and Naeem 1978) and from X-ray crystallography (Schröder et al. 2000) it became clear that the oligomer is indeed a pentamer of domain swapped dimers (D5 symmetry). Correlation of the then available TEM (Fig. 1.4) and X-ray data (Harris et al. 2001) produced a high degree of unity with a meaningful molecular superimposition (Fig. 1.5). The topic was

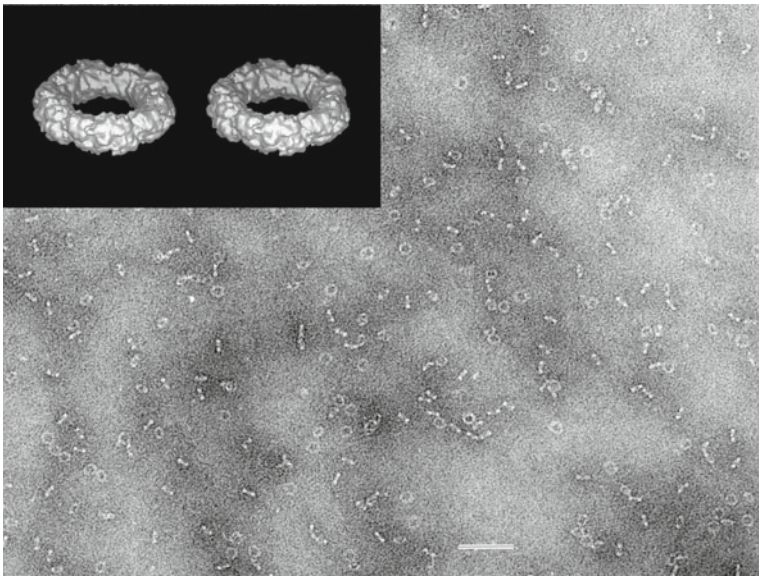

Fig. 1.4 Human erythrocyte peroxiredoxin II imaged by TEM negative staining with ammonium molybdate in the presence of trehalose (Harris et al. 2001), with a stereo pair of the 19 Å image reconstruction (inset)

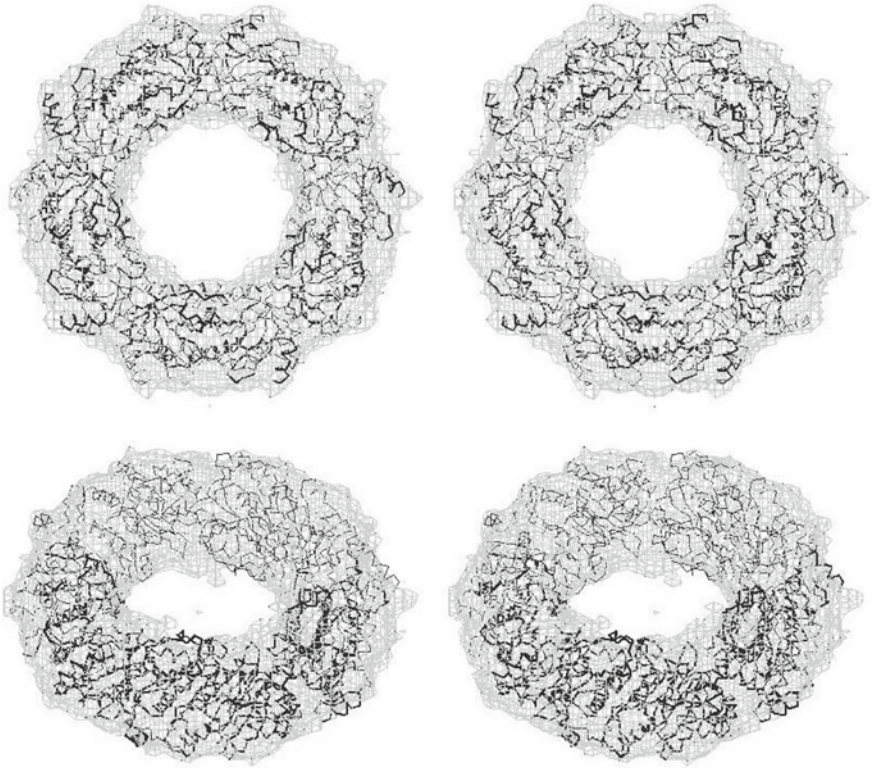

Fig. 1.5 Fitting of the 2.7 Å X-ray data of peroxiredoxin II (Schöeder et al. 2000) into the ~19 Å TEM envelope (Harris et al. 2001), as stereo pairs

reviewed in depth from a structural and functional perspective by Wood et al. (2003) and also by Cao and Lindsay (2017) in Volume 83 of this series, and Volume 44 was devoted to Peroxiredoxin Systems (Flohé and Harris 2007). Figure 1.6 depicts the historical data progression from negative stain TEM of the Prx-2 decamer to the X-ray crystal structure, within which the monomer-dimer and the dimer-pentamer interfaces have been defined (Schröder et al. 2000). Unexpectedly, when trying to produce 2-D crystals of the erythrocyte Prx-2 decamer in the presence of ammonium molybdate-PEG for TEM study, Meissner et al. (2007a) created a higher-order dodecahedron of decamers (Fig. 1.7). This assembly resembles virus-like particles and other protein cages. As a homo-oligomeric dodecahedron it emphasises the likely creation of specific non-covalent protein-protein interactions mediated by compounds used for protein crystallisation in this instance, although the creation and stabilization of this structure in conventional buffer solutions, or other crystallization conditions has yet to be demonstrated. Nevertheless, the ordered charge distribution on the peroxiredoxin decamer (Fig. 1.8) may contribute to the likelihood of dodecahedron creation. Interest in the

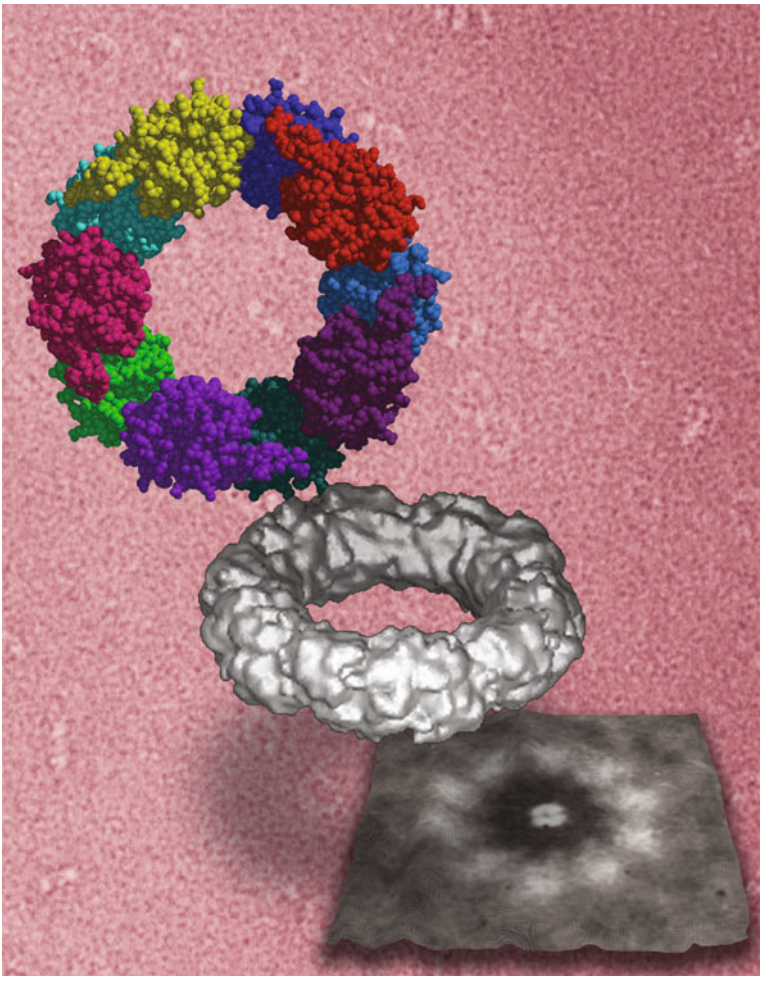

Fig. 1.6 A composite showing the first EM definition of 10 subunits with the Prx II toroid Å (Harris 1969), the Prx II molecular reconstruction from TEM| (Harris et al. 2001) and the 2.7 Å Prx II X-ray model (Schröder et al. 2000), with a background TEM image. (Courtesy of Zachary A Wood)

formation of cage-like protein assemblies and virus-like particles is expanding, revealing by TEM structures similar to the peroxiredoxin II decamer (Bale et al. 2016). Similar considerations undoubtedly apply during the formation of tubular stacks and 2D crystals of Prx II (Harris et al. 2001) and tubules of Prx III (Yewdall et al. 2018a).

Prx III is a ring of six dimers and has been shown to create double/catenated dodecamers (Cao et al. 2005, 2015). The ready formation of stacked ring tubules (nanotubules), a characteristic of Prx III, is a feature expanded upon by Phillips

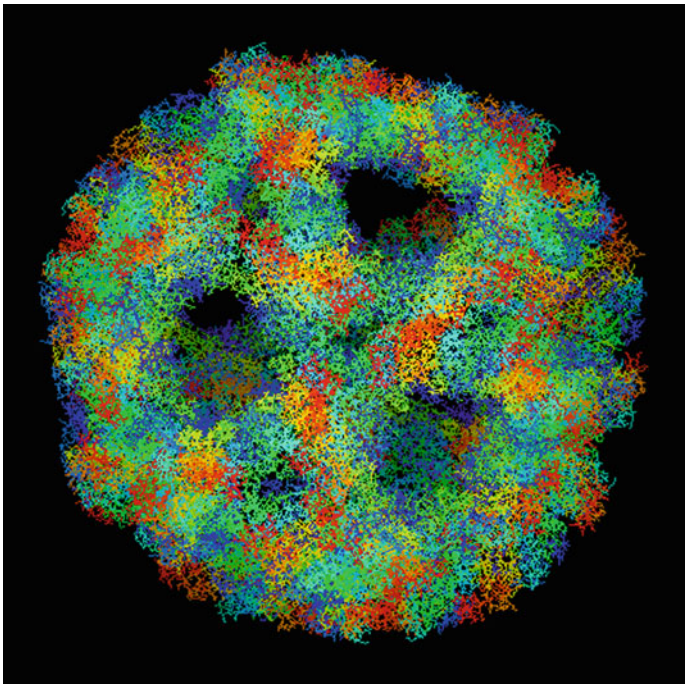

Fig. 1.7 The peroxiredoxin II dodecahedral higher-order assembly (i.e. containing 12 decamers) viewed down the three-fold axis (see Meissner et al. 2007a, b)

et al. (2014) and subsequently within the sphere of nanotechnology by Yewdall et al. (2018b) and Manuguri et al. (2018). Using the recently developed Volta phase plate has found application for the structural study of Prx-III (Khoshouei et al. 2016), with the claim that there is benefit for the study of smaller protein molecules.

Other members of the peroxiredoxin family have also received the attention of electron microscopists and more extensively X-ray crystallographers. The PDB contains many examples of peroxiredoxins, monomers, dimers and oligomers, in oxidised and reduced states and molecules from mutant genes. Thus, it is probable that these proteins and their oligomers will continue to be of structural interest.

Collagen Assemblies

Collagen Fibrils and SLS Crystallites

The fibrous proteins present special classes of protein association, in some cases bordering upon linear crystallization. Although not included in the present volume, this group of proteins were discussed in detail in Volume 82 of the series, edited by

David Parry and John Squire. Here we will consider briefly only fibrillar collagen, as a special example of protein self-association. (Other extracellular fibrils, such as the intermediate filaments, elastin, amyloid, and intracellular fibrils and tubules, including actin, myosin, titin, tropomyosin, microtubules and others, have also received the attention of electron microscopists and X-ray crystallographers. Perhaps most significant is the work on skeletal muscle myofibrils and the sliding filament mechanism of contraction (Huxley and Hanson 1953), induced by the action potential, calcium entry and control of ATP/energy utilization).

The collagen molecule is an elongated heterotrimer (~300 nm triple-helix) hydroxyproline-rich structural protein; most collagen family members possesses unique self-assembly properties in vivo and in vitro in physiological and experimental solutions, generating quasi-crystalline banded fibres aligned as bundles that possess inherent longitudinal strength and flexibility (Holmes and Kadler 2006; Harris and Reiber 2007; Harris et al. 2013; Harris 2017). The imaging and modelling of collagen fibril assembly has been considered extensively since the early days of transmission electron microscopy through to the present day, resulting in the widely accepted linear gap (~40 nm) and partial overlap (~27 nm) model with lateral alignment of heterotrimers that creates the ~67 nm "D-banded" repeating structure (Fig. 1.9). Detailed molecular aspects of the fibrillar collagens were presented by Bella and Hulmes (2017).

What is less widely appreciated is the fact that under in vitro experimental conditions the collagen molecule can also create a deviant collagen assembly, as discrete rod-like structures termed segment long spacing (SLS) bundles or *crystallites* (Fig. 1.10), containing many parallel N-/C-terminus laterally aligned

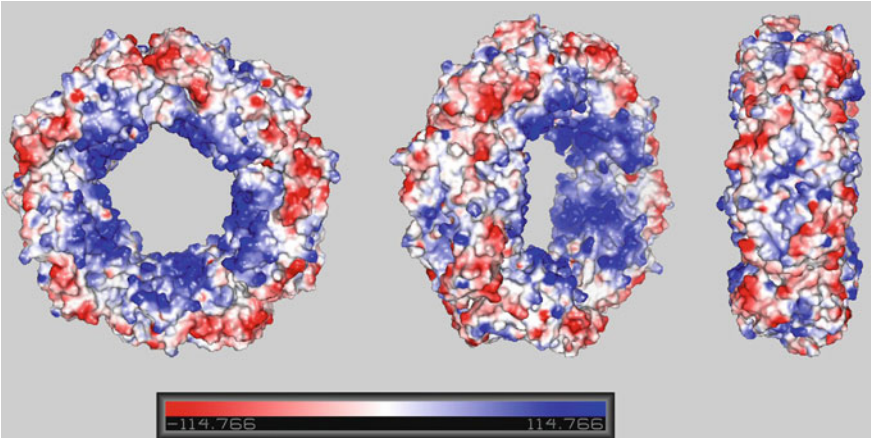

Fig. 1.8 The charge distribution on the surface of the peroxiredoxin II decamer (red negatively charged residues, blue positive residues). Note the central predominance of positive surface charge and the repeating distribution of positive and negative charge on the outer surface, available for possible inter-decamer linkage

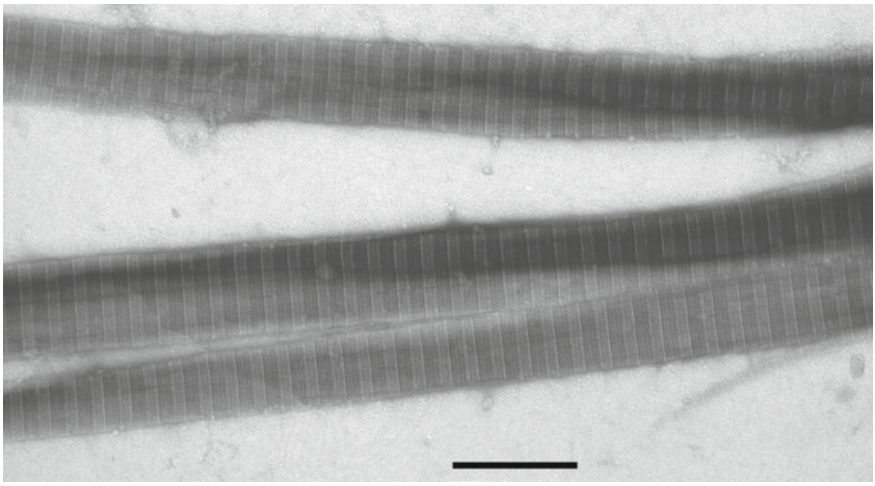

Fig. 1.9 Collagen type I fibres produced by in vitro assembly from the acid-soluble heterotrimer (Harris 2017). Each fibre contains a number of thinner fibres aligned with respect to their underlying molecular 67 nm repeating pattern of linear molecules (i.e. the established gap/overlap model). The collagen fibres are negatively stained with uranyl acetate. The scale bar indicates 400 nm

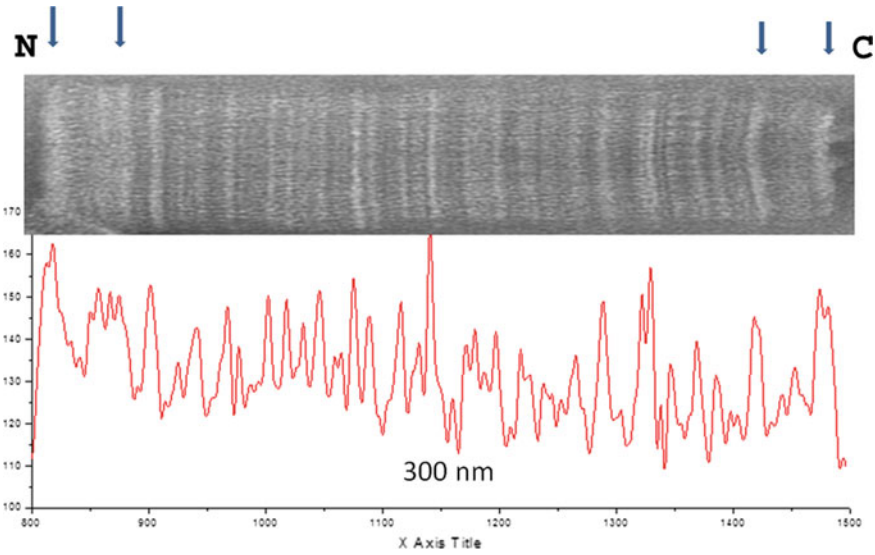

Fig. 1.10 A single collagen type I segment long spacing (SLS) crystallite (length ~300 nm), negatively stained with uranyl acetate (top). A densitometric scan (bottom) reveals the zones of increased protein thickness, indicative of the varying thickness along the individual aligned heterotromers oligomerized within the SLS bundle

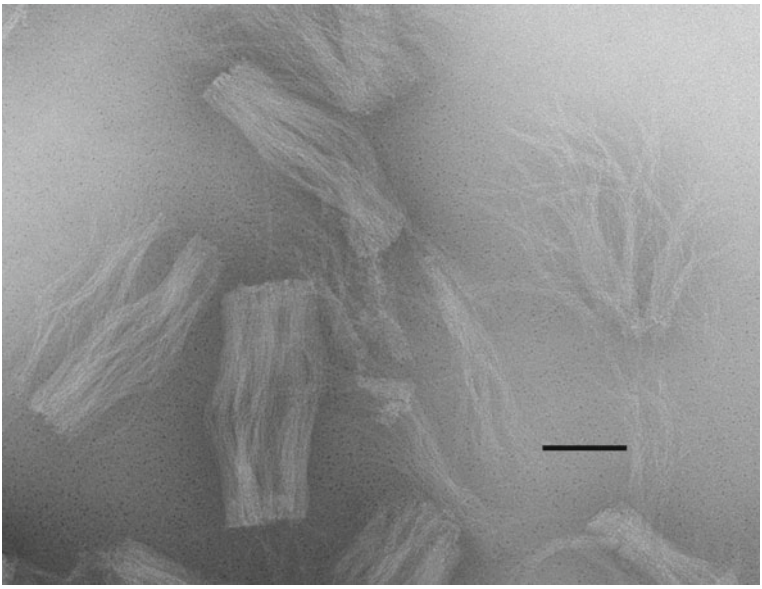

Fig. 1.11 SLS crystallites negatively stained in ammonium molybdate for TEM study undergoing spontaneous dissociation during specimen preparation. The individual hetetrotrimers can be seen to splay out from the unstable SLS bundles that have lost their characteristic banding pattern (Harris and Lewis 2016). From the original acidic collagen solution the heterotrimers have been imaged as flexible rather than straight elongated molecules (Harris et al. 2013) (The transfer to neutral pH saline solutions imparts structural linearity/rigidity during fibril formation, also during mildly acidic ATP conditions that generate SLS crystallites)

heterotrimers. (For the other deviant fibrous form of collagen, termed fibrous long spacing (FLS), there is only a $\sim$27 nm molecular overlap, with no linear spacing; see Doyle et al. 1975). The band pattern exhibited by collagen type I SLS crystallite indicates the summated variation in protein thickness along its length (densitometric scan, Fig. 1.9), in turn indicating the periodic thickness variation along a single triple helix. It is this feature of the collagen molecule that underlies and creates the more complex banding pattern exhibited by the collagen fibril. The stability of the collagen SLS crystallite, formed under mildly acidic conditions in the presence of ATP is not great, indicated by the partial dissociation back to individual hetereotrimers when negatively stained with neutral pH ammonium molybdate (Fig. 1.11) (see Harris and Lewis 2016). However, negative staining with uranyl acetate stabilizes the SLS structures, although stability can be imparted by glutaraldehyde crosslinking when negative staining with ammonium molybdate. The subtleties of fibrous protein-protein interactions undoubtedly play numerous important roles in filament creation.

High-Resolution Studies of Model Proteins

With the now widespread installation of highly-automated 300 kV cryo-electron microscopes with direct electron detectors, much recent interest has focused on pushing the effective resolution of single-particle reconstructions of protein complexes and with this, a need for well-behaved and readily available protein complex samples has grown. In an analogous manner to the utility that lysozyme and thaumatin have found for testing synchrotron beamlines and X-ray diffraction processing software, ferritin and beta-galastosidase have become the mainstay test protein complexes for cryo-electron microscopy. The *Escherichia coli* beta-galactosidase complex fulfils this role, weighing in at nearly 500 kDa the homotetramer, with its mixed alpha- and beta- secondary structure and D2 symmetry is widely used to test new microscope installations and software. A recent reconstruction using Relion-3.0 achieved a resolution of 1.9 Å (Zivanov et al. 2018). The four-helix bundle of ferritin forms a highly ordered and stable 24-meric nanocage of around half a megadalton. Horse spleen ferritin is readily available from chemical suppliers and recombinant ferritins are simple to purify, as they are usually thermostable. Using Relion, a 1.65 Å resolution reconstruction was produced of the human apo-ferritin complex, which is higher than many ferritin structures determined by X-ray crystallography. These records for published high resolution cryo-EM single particle reconstructions have already been broken on test data collected and current anecdotally reported resolutions are close to what an X-ray crystsllographer would consider to be a true 'atomic resolution' structure. We will certainly see many more single particle reconstructions at sub-2 Å resolution as new microscopes and detectors are introduced in laboratories across the world. Much like the crystallographic test samples, lysozyme and thaumatin, our knowledge of the biochemistry and structure of ferritin and beta-galactosidase has not been significantly advanced by these new high-resolution reconstructions. They do however provide an important benchmark and show what can be done by modern microscopes, detectors, and software (Fig. 1.12).

Encapsulins

The encapsulins are a class of protein complexes that are of growing interest, from both a basic biological perspective and as tools in applied biotechnology. The encapsulins were originally identified as a 30 nm diameter virus-like particles in the hyperthermophilic archaeaon *Pyrococcus furiosus* (Namba et al. 2005), a later crystal structure of these particles highlighted their structural relationship to bacteriophage capsids (Tatur et al. 2007). An accidental discovery of a related virus-like particle in *Thermotoga maritima*, with an encapsulated ferritin-like protein led to the coining of the term 'encapsulin' for the proteins forming these bacterial nanocompartments (Sutter et al. 2008). Close analysis of the crystal

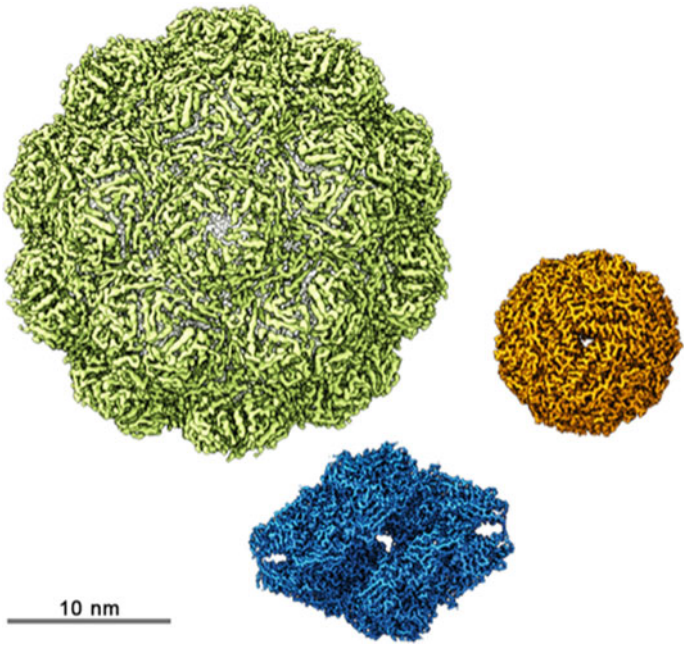

Fig. 1.12 Depictions of the electron potential maps of some recent high-resolution single-particle reconstructions of protein complexes. The T = 3 icosahedral *Myxococcus xanthus* encapsulin is shown in lime yellow (EMD: 5917) (McHugh et al. 2014); octahedral horse spleen ferritin in orange (EMD: 0263) (Zivanov et al. 2018); and the tetrameric D2-symmetry *E. coli* beta-galactosidase in blue (EMD: 0153) (Zivanov et al. 2018). The scale bar represents 10 nm

structure of the *T. maritima* encapsulin showed a short peptide sequence from the encapsulated enzyme bound to the interior wall of the nanocage. This peptide is responsible for directing enzyme cargoes to the encapsulin cage and it has been shown to be sufficient to direct heterologous proteins to recombinant encapsulins (Tamura et al. 2014).

Interest in encapsulins intensified after 2013, when primary research began appearing more frequently on the structure and function of these fascinating complexes. Since then, a single-particle cryo-EM reconstruction of the *Myxococcus xanthus* encapsulin has been published (Fig. 1.12) (McHugh et al. 2014), along with a detailed study of the biochemistry of the new family of ferritin-like proteins found within many of these proteins (He et al. 2016). What is striking about these encapsulated-ferritins is their annular decameric structure, which is in contrast to the 24-meric nanocages formed by the classical ferritins. The encapsulated ferritins function as ferroxidase enzymes to oxidise iron(II) to iron(III) within the encapsulin nanocage, where it is stored as inert ferrihydrite, and iron-phosphate minerals. What is striking about these complexes, is their ability to hold up to ten times more iron than the classical ferritin nanocages (He et al. 2016; McHugh et al. 2014). The role that such a massive iron store plays in bacteria is still open to speculation. A recent

survey of bacterial and archaeal species highlighted the widespread distribution of encapsulins across diverse environmental niches and identified a number of new cargo proteins found within them, including new ferroxidase enzymes found only in bacteria of the firmicutes phylum, and enzymes involved in the response to nitrosative stress (Giessen and Silver 2017). It is clear from this survey that the primary role of encapsulins is to either protect the host cell from oxidative damage caused by the redox reactions carried out by their enzyme cargoes, or to protect the host from oxidative damage caused by the substrates of these cargo enzymes.

The ability to target non-native proteins to the interior of the encapsulin cage, coupled with its stability and amenability to surface modification, has led to the adoption of encapsulins for biotechnological applications that range from the targeted delivery of drugs (Moon et al. 2014), to imaging tools for studying mammalian cells (Sigmund et al. 2018). While these applications are still in the early stages of development and implementation in medicine and biotechnology, there is still much to be learnt about the basic biology and function of these complexes in their host organisms.

This introductory chapter cannot cover all interesting protein complexes and represents only a selection of specific interest to the editors of this volume. Indeed, in this series of books, we can only hope to cover a small selection of the recent advances in the study of macromolecular complexes. The importance of higher-order structure in biology and the coming together of multiple proteins, is central to the function of all organisms. As our technology for the high-resolution study of proteins advances, in particular our ability to image and generate 3D reconstructions of proteins by cryo-electron microscopy, more interesting macromolecular complex structures are being elucidated and their functions illuminated through the combination of structural and functional data. We hope that these selected chapters are of interest to our readers and that authors whose work we have not been able to cover in this introduction, and who have not contributed to this volume, are not offended by the omission of their work. Finally, we would like to thank all of our contributors, who have all provided interesting considerations of their work and its context within the wider biochemistry and biology of their host systems and organisms.

References

Adrian M, Dubochet J, Lepault J, McDowall AW (1984) Cryo-electron microscopy of viruses. Nature 308:32–36

Afanasyev P, Seer-Linnemayr C, Ravelli RBG, Matadeen R, De Carlo S, Alewijnse B, Rodrigo V. Portugal RV, Pannu NS, Schatz M, van Heel M (2017) Single-particle cryo-EM using alignment by classification (ABC): the structure of *Lumbricus terrestris* haemoglobin. IUCrJ 4: 678–694

Ahnert SE, Marsh JA, Hernández H, Robinson CV, Teichmann SA (2015) Principles of assembly reveal a periodic table of protein complexes. Science 350. https://doi.org/10.1126/science.aaa2245

Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, Baker D (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. Science 353:389–394

Bella J, Hulmes DJS (2017) Fibrillar collagens. In: Parry DAD, Squire JM (eds) Fibrous proteins: structures and mechanisms. Subcellular biochemistry vol. 82. Springer Nature, pp 457–490m

Bernal JD, Crowfoot D (1934) X-ray photographs of crystalline pepsin. Nature 133:794–795

Borhani HA, Berghmans H, Trashin S, DeWael K, Fago A, Moens L, Habibi-Rezaei M, Dewilde S (2012) Kinetic properties and heme pocket structure of two domains of the polymeric hemoglobin of Artemia in comparison with the native molecule. Biochim Biophys Acta 1854:1307–1316

Cao Z, Roszak AW, Gourlay LJ, Lindsay JG, Isaacs NW (2005) Bovine mitochondrial peroxiredoxin III forms a two-ring catenane. Structure 13:1661–1664

Cao Z, McGow DP, Shepherd C, Lindsay JG (2015) Improved catenated structures of bovine Peroxiredoxin III F190L reveal details of ring-ring interactions and a novel conformational state. PLoS ONE 10(4):e0123303. https://doi.org/10.1371/journal.pone.0123303

Cao Z, Lindsay JG (2017) The peroxiredoxin family: an unfolding story. In: Harris JR, Marles-Wright J (eds) Macromolecular protein complexes. Springer Nature, pp 127–147

Chen W-T, Chen Y-C, Liou H-H, Chao C-Y (2015) Structural basis for cooperative oxygen binding and bracelet-assisted assembly of *Lumbricus terrestris* haemoglobin. Scientific Reports 5:9494. https://doi.org/10.1038/srep09494

De Carlo S, Harris JR (2011) Negative staining and cryo-negative staining of macromolecules and viruses for TEM. Micron 42:117–131

Doyle BB, Hukins DWL, Hulmes DJS, Miller A, Woodhead-Galloway J (1975) Collagen polymorphism: its origins in the amino acid sequence. J Mol Biol 91:79–99

Fitzpatrick AWP, Falcon, He S, Murzin AG, Murshudov G, Garringer HJ,- RA, Ghetti B, Goedert M, Scheres SHW (2017) Cryo-EM structures of Tau filaments from Alzheimer's disease brain. Nature. 547(7662):185–190. https://doi.org/10.1038/nature23002

Flohé L, Harris JR (2007) Peroxiredoxin systems, subcellular biochemistry, vol. 44. Springer, New York

Gatsogiannis C, Hofnagel O, Markl J, Raunser S (2015) Structure of mega-hemocyanin reveals protein origami in snails. Structure 23:93–103

Gia Z, Matsuno A, Kato K, Kato S, Khan MRI, Shimizu T,Yoshioka T,Kato Y, Kishimura H, Kanno G, Miyabe Y, Terada T, Tanaka Y, Yao M (2015) Crystal structure of the 3.8-MDa respiratory supermolecule hemocyanin at 3.0 Å. Resolution Structure 23:2204–2212

Giessen TW, Silver PA (2017) Widespread distribution of encapsulin nanocompartments reveals functional diversity. Nature Microbiol 2 Article number: 17029. https://doi.org/10.1038/nmicrobiol.2017.29

Grabowski M, Niedzialkowska E, Minor W (2016) The impact of structural genomics: the first quindecennial. J Struct Funct Genomics 17:1–16

Green NM (1972) Analysis of the Structure of Complex Proteins by Electron Microscopy. In: Jaenicke R, Helmreich E (eds) Protein-protein interactions., Springer| Verlag, pp 183–211

Green MN, Valentine R, Wrigley NG, Ahmad F, Jacobson B, Wood HG (1972) Transcarboxylase XI. Electron microscopy and subunit structure. J Biol Chem 247:6284–6298

Green NM (1990) Avidin and streptavidin. Methods Enzymol 84:51–67

Harding MM, Crowfoot Hodgkin D, Kennedy AF (1966) The crystal structure of insulin ll. An investigation of rhombohedral zinc insulin crystals and a report of other crystalline forms. J Mol Biol 16:212–226

Harris JR (1968) Release of a macromolecular protein component from human erythrocyte ghosts. Biochim Biophys Acta 150:534–537

Harris JR (1969) The isolation and purification of a macromolecular protein component from the human erythrocyte ghost. Biochim Biophys Acta 188:31–42

Harris JR (2015) Transmission electron microscopy in molecular structural biology: a historical survey. Arch Biochem Biophys 581:3–18

Harris JR (2017) Visualizing in vitro type I collagen fibrillogenesis by transmission electron microscopy. In: Rittié L (ed) Fibrosis: methods and protocols. Methods in Molecular Biology, vol. 1627. Springer Science + Business Media LLC, pp 367–383

Harris JR, De Carlo S (2014) Negative staining and Cryonegative staining. In: John Kuo (ed.), Electron microscopy: methods and protocols. Methods in Molecular Biology, vol. 1117. © Springer Science + Business Media, New York, pp 215–258. https://doi.org/10.1007/978-1-62703-776-1_11

Harris JR, Naeem I (1978) The subunit composition of two high molecular weight extrinsic proteins from human erythrocyte membranes. Biochim Biophys Acta 537:495–500

Harris JR, Schroder E, Isupov MN, Scheffler D, Kristensen P, Littlechild JA, Vagin AA, Meissner U (2001) Comparison of the decameric structure of peroxiredoxin II by transmission electron microscopy and X-ray crystallography. Biochim Biophys Acta 1547:221–234

Harris JR, Meissner U, Gebauer W, Markl J (2004) 3D reconstruction of the hemocyanin subunit dimer from the chiton *Acanthochiton fascicularis*. Micron 35:23–26

Harris JR, Lewis RJ (2016) The collagen type I segment long spacing (SLS) and fibrillar forms: formation by ATP and sulphonated diazo dyes. Micron 86:36–47

Harris JR, Reiber A (2007) Influence of saline and pH on collagen type I fibrillogenesis in vitro: fibrin polymorphism and colloidal gold binding. Micron 38:115–521

Harris JR, Soliakov A, Lewis RJ (2013) In vitro fibrillogenesis of collagen type I in varying ionic and pH conditions. Micron 49:60–68

He D, Hughes S, Vanden-Hehjir S, Georgiev A, Altenbach K,Tarrant E, MackaY cl, Waldron KJ, Clarke DJ (2016) Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. eLIFE. https://doi.org/10.7554/eLife.18972.001

Henderson R (2015) Overview and future of single particle electron cryomicroscopy. Arch Biochem Biophys 581:19–24

Holmes DF, Kadler KE (2006) Collagen Fibril assembly in vitro. In: Harris R, Graham J, Rickwood D (eds) Cell biology protocols. Wiley, pp 375–378

Huxley HE, Hanson EJ (1953) Structural basis of the cross-striations in muscle. Nature 172: 530–532

Iadanza MG, Silvers R, Boardman J, Smith HI, Karamanos TK, Deb elouchina GT, Su Y, Griffin RG, Ranson NA, Radford SE (2018) The structure of a β_2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. Nat Commun. https://doi.org/10.1038/s41467-018-06761-6

Ji X, Axford D, Owen R, Evans G, Ginn HM, Sutton G, Stuart DI (2015) Polyhedra structures and the evolution of the insect viruses. J Struct Biol 192:88–99

Kato S, Matsui T, Gatsogiannis C, Tanaka Y (2017) Molluscan hemocyanin: structure, evolution, and physiology Biophys Rev. https://doi.org/10.1007/s12551-017-0349-4

Kendrew JC, Dickerson RE, Hart RG, Davies DR, Phillips DC, Shore VG (1960) Structure of myoglobin: a three dimensional Fourier synthesis at 2 Å. Nature 185:422–427

Khoshouei M, Radjainia M, Phillips AJ, Gerrard JA, Mitra AK, Plitzko JM, Baumeister W, Danev R(2016) Volta phase plate cryo-EM of the small protein complex Prx3. Nature. https://doi.org/10.1038/ncomms10534|www.nature.com/naturecommunications

Laemli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Li J, Carroll J, Ellar DJ (1991) Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution Nature 353:815–821

Manuguri V, Webster K, Yewdall NA, An Y, Venugopal H, Bhugra V, Turner A, Domigan LJ, Gerrard JA, Williams DE, Malamoström (2018) Assembly of protein stacks with in situ synthesized nanoparticle cargo. Nano Lett. 18:5138–5145. https://doi.org/10.1021/acs.nanolett.8b02055

Markl J (2013) Evolution of molluscan hemocyanin structures. Biochim Biophys Acta 1834:1840–1852

Martin AG, Depoix F, Stohr M, Meissner U, Hagner-Holler S, Hammouti K, Burmester T, Heyd J, Wriggers W, Markl J (2007) Limulus polyphemus Hemocyanin: 10 Å Cryo-EM structure,

sequence analysis, molecular modelling and rigid-body fitting reveal the interfaces between the eight hexamers. J Mol Biol 366:1332–1350
Marsh JA, Teichmann SA (2014) Structure, dynamics, assembly, and evolution of protein complexes. Ann Rev Biochem https://doi.org/10.1146/annurev-biochem-060614-034142
Matthews BW, Bernhard SA (1973) Structure and symmetry of oligomeric enzymes. Ann Rev Biophys Bioeng 2:257–317
McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Heymann JB, Winkler DC, Lam AS, Wall JS, Steven AC, Hoiczyk E (2014) A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. EMBO J 33:1891–1917
McMullan G, Faruqi AR, Henderson R (2016) Direct electron detectors. Meth Enzymol 579:1–16
Meissner U, Gatsogiannis C, Moeller A, Depoix F, Harris JR, Markl J (2007a) Comparative 11 Å structure of two molluscan hemocyanins from 3D cryo-electron microscopy. Micron 38:754–768
Meissner U, Schröder E, Scheffler D, Martin AG, Harris JR (2007b) Formation, TEM study and 3D reconstruction n of the human erythrocyte peroxiredoxin-2 dodecahedral higher-order assembly. Micron 38:29–39
Moon H, Lee J, Min J, Kang S (2014) Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform. Biomacromol 15:3794–3801
Namba K, Hagiwara K, Tanaka H, Nakaishji Y (2005) Expression and molecular characterization of spherical particles derived from the genome of the hyperthermophilic euryarchaeote *Pyrococcus furiosus*. J Biochem 138:193–199
Nicholls RA, Tykac M, Kovalevskiy O, Murshudov GN (2018) Current approaches for the fitting and refinement of atomic models into cryo-EM maps using CCP-E. Acta Crystallog D74 https://doi.org/10.1107/s2059798318007313
Oliver RM (1973) Negative stain electron microscopy of protein macromolecules. Meth Enzymol 27:617–672
Owen RL, Juanhuix J, Fuchs M (2016) Current advances in synchrotron radiation instrumentation for MX experiments. Arch Biochem Biophys 602:21–31
Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature 185:416–422
Phillips AJ, Littlejohn J, Yewdall NA, Zhu T, Valéry C, Pearce FG, Mitra AK, Radjainia M, Gerrard JA (2014) Peroxiredoxin is a versatile self-assembling tecton for protein nanotechnology. Biomacromol 15:1871–1881
Rodis P, Hoff JE (1984) Naturally occurring protein crystals in the potato. Plant Physiol 74:907–911
Rousselot T, Jaenicke E, Lamkemeyer T, Harris JR, Pirow R (2006) Native and subunit molecular mass and quarternary structure of the hemoglobin from the primitive branchiopod crustacean *Triops cancriformis.* Febs J. https://doi.org/10.1111/j.1742-4658.2006.05408.x
Royer WE, Kristen Strand K, van Heel Hendrickson WA (2000) Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. Proc Natl Acad Sci 97:7107–7111
Ryle AP, Sanger F, Smith LF, Kitai R (1955) The disulphide bonds of insulin. Biochem J 60:541–556
Schroder E, Littlechild JA, Lebedev AA, Errington N, Vagin AA, Isupov MN (2000) Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 A resolution. Structure 8:605–615
Sigmund F, Massner C, Erdmann P, Stelzl A, Rolbieski H, Desai M, Bricault S, Wörner TP, Snijder J, Geerlof A, Fuchs H, Hrabě de Angelis M, Heck AJR, Jasanoff A, Ntziachristos V, Plitzko J, Westmeyer GG (2018) Bacterial encapsulins as orthogonal compartments for mammalian cell engineering. Nat Commun. https://doi.org/10.1038/s41467-018-04227-3
Stahlberg H, Biyani N, Engel A (2015) 3D reconstruction if two-dimensional crystals. Arch Biochem Biophys 581:68–77