Protein Quaternary Structure: Subunit–Subunit Interactions

Susan Jones, *University College, London, England*

Janet M Thornton, *University College, London, England*

The quaternary structure of proteins is the highest level of structural organization observed in these macromolecules. The multimeric proteins that result from quaternary structure formation involve the association of protein subunits through hydrophobic and electrostatic interactions. Protein quaternary structure has important implications for protein folding and function.

Introduction

Proteins are organized into a structural hierarchy. The polypeptide chain at the primary structural level comprises a linear, noncovalently linked amino acid residue sequence. Secondary structure is the level at which the linear sequences aggregate to form structural motifs such as helices and sheets. The tertiary structure is formed by packing of the secondary structural elements into one or more compact globular domains. In many cases proteins are composed of only a single polypeptide chain that has tertiary structure as its highest level of organization, e.g. lysozyme. These are termed monomeric proteins. However, many proteins are composed of more than one polypeptide chain, associated into assemblies possessing a specific quaternary structure, e.g. dimeric interleukin 8 and tetrameric $\alpha_2\beta_2$ human haemoglobin (**Figure 1**). The most complex assemblies are those observed in the higher order structures, such as the icosahedral viruses, that comprise 60 monomers in identical symmetrical positions. The quaternary structure of a protein describes the stoichiometry and stereochemistry of assemblies of noncovalently linked subunits, characterized by the lower levels of structural organization (Jaenicke, 1987).

In a discussion of protein quaternary structure it is important to adhere to a single set of definitions. Those widely used in the literature, and adopted here, were derived by Monod *et al.* (1965) in their theoretical model of allosteric effects in protein structures. An oligomer is defined as a protein assembly containing a finite, relatively small number of identical subunits. Protomers are defined as the identical subunits associated within an oligomeric protein. A monomer is defined as the fully dissociated protomer, or any protein that is not made up of subunits. A subunit is purposely undefined, and may be used to refer to any chemically or physically identifiable submolecular entity within a protein, whether identical to or different from, other components. Using these definitions, the haemoglobin tetramer (comprised of two α and two β polypeptide chains) is defined as an oligomer consisting of two protomers, each consisting of two monomers, i.e. one α and one β polypeptide chain. The definition of a subunit allows the term to be used for either the α- or β-monomer, or for the $\alpha\beta$-protomer. The term multimer is also widely used in the literature and is defined here as a protein with a finite number of subunits that need not be identical.

The quaternary nature of some proteins was first identified from centrifugation experiments devised in the 1920s to calculate the molecular weights of proteins. Since then, combinations of association and hybridization techniques have led to the discovery of a large number of proteins possessing quaternary structure. To date, the three-dimensional structures of many hundreds of multimeric proteins with identical subunits (oligomers) and nonidentical subunits have been solved by X-ray crystallography. An example of an oligomeric protein solved by this method is aspartyl protease from human immunodeficiency virus 1 (HIV-1) retrovirus (Navia *et al.*, 1989). This protein functions to release structural proteins and enzymes (such as reverse transcriptase and integrase) from viral polyprotein products. HIV-1 protease is a homodimer, in which each subunit comprises almost exclusively β sheet, turn and extended polypeptide structural elements (**Figure 1**). Each subunit contributes one highly conserved Asp-Thr-Gly catalytic triad sequence to form a symmetric active site in the dimer. An example of a multimeric protein with nonidentical subunits solved by X-ray crystallography is the glycoprotein hormone, human chorionic gonadotrophin (Lapthorn *et al.*, 1994). This protein is secreted by the placenta in the early weeks of pregnancy and stimulates the secretion of the steroid progesterone. The protein is an $\alpha\beta$-heterodimer in which each subunit has a similar extended topology that includes a cysteine knot motif, common to a number of growth factors. The α and β
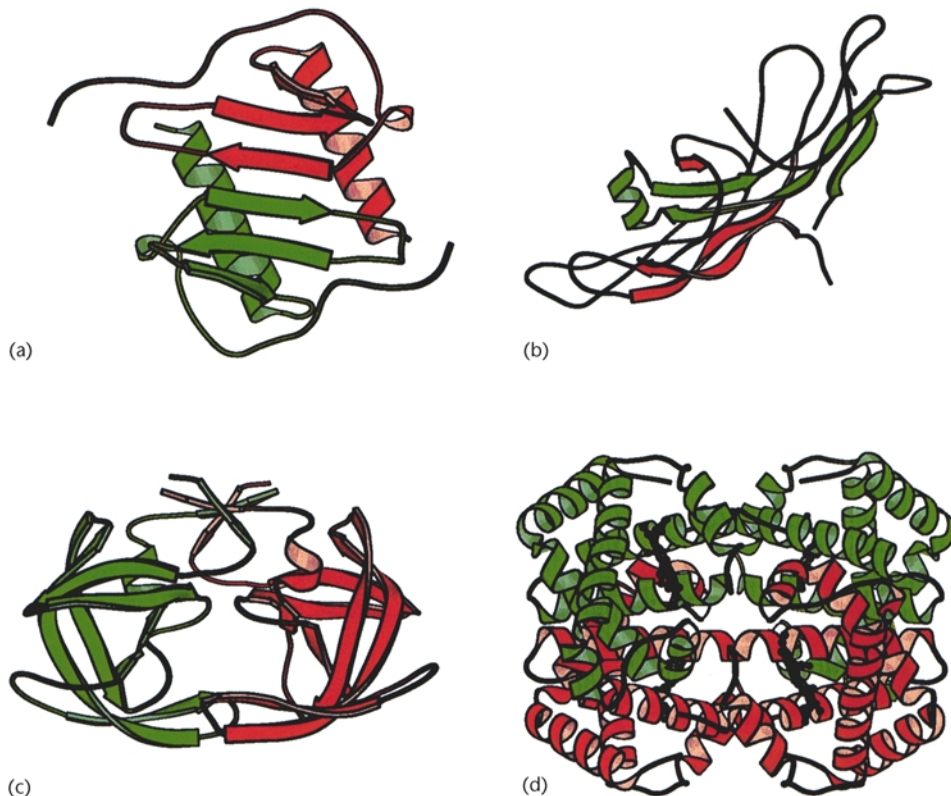
Figure 1 Molscript diagrams depicting the secondary structure elements and the quaternary structure of (a) homodimeric interleukin 8; (b) heterodimeric human chorionic gonadotrophin; (c) homodimeric HIV-1 protease; (d) heterotetrameric human haemoglobin. In each diagram the protein subunits are differentiated by colour, and in (d) one αβ-protomer of haemoglobin is coloured red and one green. The haem groups in each subunit shown in (d) are depicted by ball-and-stick representations.

subunits are integrally associated with a segment of the β subunit wrapped around the α subunit and covalently linked by a disulfide bond (**Figure 1**).

Quaternary Structure Assembly

Protein stoichiometry

The stoichiometry of proteins possessing a quaternary structure considers the number of subunits involved in the assemblies. The association of subunits in quaternary structure can lead to the formation of closed structures, of which dimers and tetramers are by far the most frequently observed. In addition, quaternary proteins can exhibit open elongated polymer structures. The way in which different numbers of protein subunits associate to form aggregates has been defined as macroassociation.

The process of macroassociation is divided into three modes, namely heterologous, isologous and pseudoisologous. Each can be defined using two terms: binding set (the residues of one protomer involved in binding to one other

protomer), and domain of bonding (the two, linked binding sets) (Monod *et al.*, 1965). In heterologous associations the domain of bonding is made up of two different binding sets, and in isologous associations the two binding sets involved are identical (**Figure 2**). In pseudoi-sologous associations, the domain of bonding comprises two almost identical binding sets. In isologous associations the binding set of each protomer is 'covered' by the equivalent binding set on the other protomer, hence these associations tend to lead to finite closed structures.
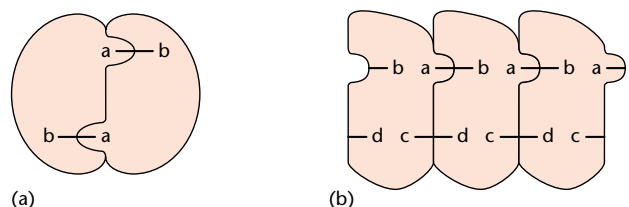

Figure 2 Modes of association in multimeric proteins: (a) isologous, in which the binding sets (indicated by the letters 'a' and 'b') are identical; (b) heterologous, in which the binding sets (indicated by the letters 'a', 'b', 'c' and 'd') are not identical. Adapted from Monod *et al.*, 1965.

Heterologous associations can involve multiple binding sets on a single protomer that can lead to infinite open structures. If an oligomer has an odd number of equivalent protomers, then the associations between them must be heterologous; however, if an oligomer has an even number of equivalent protomers, the associations can be either isologous, heterologous or a mixture of the two. Some elongated proteins, such as actin, are formed by indefinite heterologous association of globular protomers; however, heterologous associations can also lead to closed structures, such as those observed in the trimeric bacteriochlorophyll protein and the tetrameric manganese superoxide dismutase.

The definition of different modes of association raised the question of which one is most prevalent amongst proteins. Monod *et al*. (1965) proposed that the exclusive use of isologous associations would lead only to dimers and tetramers. In support of this is the prevalence of dimers and tetramers amongst the proteins solved, and by implication isologous associations. However, thermodynamic calculations on the possibilities of all-isologous, all-heterologous and mixed structures give no indication that isologous associations are more energetically favourable than heterologous.

Protein stereochemistry

Stereochemistry (the spatial arrangements of subunits within a structure) at the quaternary structure level involves the concept of symmetry. An initial understanding of the importance of symmetry in oligomeric proteins was derived principally from the comparative studies of myoglobin and haemoglobin. The importance of symmetry in terms of protein structures was also introduced in the theoretical model of the allosteric effects of enzymes (Monod *et al*., 1965). From these, and the increasing number of protein structures solved by X-ray crystallography, it was found that many subunits of oligomeric proteins (those with identical subunits) were organized into stable arrays with high symmetry. For example, in general, proteins with two identical subunits (e.g. malate dehydrogenase and triosephosphate isomerase) have their subunits arranged with twofold rotational symmetry. In a similar way, proteins with three identical subunits (e.g. bacteriochlorophyll protein and 2-keto-3-deoxy-6-phosphogluconic aldolase) generally have their subunits arranged with threefold rotational symmetry. Tetrameric proteins (e.g. glyceraldehyde-3-phosphate dehydrogenase and lactate dehydrogenase) commonly exhibit dihedral 222 symmetry. There are, however, oligomeric proteins that do not exhibit symmetry at the quaternary level of structure. One example of this is observed in yeast hexokinase. This structure contains two identical subunits related by a rotational and a transitional symmetry element (not the expected 180° rotation). In addition it is possible that in some structures the association itself may be symmetrical but minor structural changes between the subunits may exist. Such changes generally only exist in the crystal form of the protein, where identical, symmetrically related subunits might be in an anisotropic crystal environment.

The symmetry of an oligomeric protein affects its properties and functions, and can be crystallographic or noncrystallographic. The unit cell of a crystal is the basic building block, repeated infinitely by translation in three dimensions. The asymmetric unit is the basic repeating unit, which is related to all the other identical units in the unit cell by the operation of the symmetry elements. If the volume of the asymmetric unit of a crystal accommodates just one subunit of an oligomeric protein, then the other subunit(s) will be related to it by the same symmetry operation(s) that relate the asymmetric units to each other. In this case the symmetry of the oligomer is expressed in the crystallographic symmetry. Alternatively, the asymmetric unit of the crystal may accommodate the whole oligomer or more than one oligomer. In this case the symmetry of the protein will not be determined by the crystallographic symmetry but by a noncrystallographic symmetry operation.

All proteins exhibiting quaternary structure can be termed biological complexes, in that they are associations known to exist in solution and hence, by inference, in the cell. A completely different set of protein–protein complexes are those represented by associations observed in crystal packing, termed crystallographic complexes. The problem of distinguishing between crystallographic complexes and true biological complexes is a difficult one. Protein–protein interactions in crystal packing differ significantly from biological complexes. Crystal packing contacts have no biological role and hence are not subject to evolutionary pressures.

Folding and Function

The subunits of multimeric proteins can be considered as independent folding units. In these structures, protein folding probably begins with the folding of the independent subunits (the same as in monomeric proteins) and continues until the formation of a specific recognition site that can be identified by another monomer. At this stage the folding pathway shifts from being intramolecular to intermolecular, to yield a dimer structure. The dimeric structure may then undergo further folding steps to form a native protein or a folding intermediate, with a further recognition site that permits a second association step to occur. In this way the folding of multimeric structures is a succession of monomolecular folding steps and bimolecular association steps (Jaenicke, 1987). The high specificity of the association step is fundamental to the correct folding of multimeric proteins.

Multimeric proteins are, in general, functionally more versatile than proteins comprising a single polypeptide chain. It is important to consider the difference between the sum of the isolated subunits and the complete monomer. In multienzyme complexes individual subunits catalyse distinct consecutive reactions. In such structures the proximity of the reaction sites leads to enhanced activity by providing higher local concentrations of substrate about the active site of the subunit catalysing the second reaction. An example of this is observed in tryptophan synthetase, an enzyme that catalyses the final reactions of tryptophan biosynthesis. The $\alpha_2\beta_2$-multimer is in equilibrium with its constituent α and β subunits. The subunits show two distinct functions: the α subunit catalyses the conversion of indole-glycerol phosphate to indole, while the β subunit (usually present as a dimer β_2) catalyses the formation of tryptophan from indole. The complete multimer exhibits a higher rate of the reactions than the isolated subunits.

Also significant are protein functions that result from intersubunit contacts, such as allosteric interactions. Aspartate transcarbamoylase in *Escherichia coli* is a well-studied allosteric enzyme that shows cooperative effects in substrate binding and is subject to feedback control. The quaternary structure of this enzyme includes two types of subunit, a catalytic subunit and a regulatory subunit. In the native enzyme there are two catalytic subunits and three regulatory subunits present, and the cooperative mechanism is achieved through the interaction between one pair of catalytic subunits connected by a bridging regulatory subunit.

As well as catalytic functions, quaternary structure may also serve to confer additional stability to protein structures. It is also possible that proteins form quaternary structures in certain conditions to avoid an excessive osmotic pressure. Another possible function is that a greater size of macromolecules may be important in compartmentalization within the cell or in protein turnover. The association of protein subunits is also the basis of some common diseases. In sickle cell anaemia a single mutation in the β subunit of haemoglobin causes the deoxygenated form of haemoglobin to polymerize into long fibres. Alzheimer disease is characterized by the association of β-amyloid proteins to form brain lesions termed senile plaques. These last two examples illustrate how protein associations can result in adverse as well as advantageous effects in the cell.

Protein–Protein Recognition Sites

Protein–protein associations that occur in the quaternary structure formation involve the specific complementary recognition of two macromolecules to form a stable assembly. The recognition process involves factors favouring and opposing the stable association. Hydrophobic and electrostatic interactions favour the association. The loss of translational and rotational freedom of amino acids on binding opposes the association. The affinity for two molecules is determined by the change in energy and entropy of a system that contains the two proteins and solvent and the complex and solvent; however, the lack of experimental binding-association data has meant that the relative contributions of factors contributing to the binding energy of association remains unclear.

Hydrophobic interactions

The hydrophobic interaction is considered to be the primary driving force in the stabilization of protein associations. The term 'hydrophobic interaction' is used to describe the gain in free energy upon the association of nonpolar residues of proteins in an aqueous environment. The process of folding and protein–protein aggregation reduces the surface of a protein in contact with water. This is the structural basis of the hydrophobic effect in proteins. The folding of polypeptide chains and aggregation of subunits buries the hydrophobic residues of the proteins, and hence minimizes the number of thermodynamically unfavourable solute–solvent interactions. The quantitative evaluation of exactly how much hydrophobic interactions contribute to the stabilization of protein–protein associations is controversial. The controversy is based on different definitions and interpretations of the hydrophobic effect in proteins. Empirical calculations have led to energy values of between, 25 and 72 calories per $\mathring{A}^2$ ($1 \mathring{A}^2 = 0.01 \, nm^2$) of accessible surface area gained on association. These energy values are important when considering the minimum size of recognition sites in multimeric proteins and other protein–protein complexes.

Electrostatic interactions

Electrostatic interactions, in addition to hydrogen bonds and van der Waals interactions, are considered of secondary importance in protein associations (Chothia and Janin, 1975); however, the hydrogen bond (a polar interaction between donor and acceptor electronegative atoms) is an intrinsic component of protein–protein interactions. Hydrogen bonds between protein molecules are more favourable than those made with water, and hence intermolecular hydrogen bonds contribute to the binding energy of association. It has been proposed that whereas hydrophobic forces drive protein–protein interactions, hydrogen bonds and salt bridges confer specificity (Fersht, 1987). It has been observed that the geometry of hydrogen bonds across protein–protein interfaces (such as those in multimeric proteins) are generally less optimal and have a wider distribution than those observed in the interior of proteins. This leads to the proposal that intermolecular hydrogen bonds are weaker than those in

protein interiors. Salt bridges across the binding interface of multimeric structures can also significantly enhance stability in some complexes.

Van der Waals interactions occur between all neighbouring atoms, but those interactions at the interface are not more energetically favourable than those made with the solvent; however, they are more numerous, as the tightly packed interfaces are more dense than the solvent. Hence these interactions also contribute to the binding energy of association.

Shape complementarity

The complementarity of protein interfaces is derived from both electrostatic interactions and shape. Shape complementarity has been characterized by the size of the buried surface, and the packing density of interface atoms. Many methods (Chothia and Janin, 1975; Lawrence and Colman, 1993; Jones and Thornton, 1996) have been employed to measure packing of protein–protein subunits, with the general conclusion that subunits in multimeric proteins are tightly packed. Such methods have also revealed that there are differences in interface packing between different types of protein–protein complex. Proteins exhibiting quaternary structure have protein–protein interfaces that are more closely packed than other types of protein–protein associations, such as enzyme–inhibitor and antibody–antigen complexes. Such differences possibly reflect the evolutionary time scale of these structures (Jones and Thornton, 1996).

Recognition site properties

The number and type of interactions in multimeric proteins are generally considered in relation to the area of the subunit interface. This can be measured in terms of accessible surface area (ASA). The native structure of proteins exists only in the presence of water, and the ASA describes the extent to which protein atoms can form contacts with water. Lee and Richards (1971) were the first to propose the concept of ASA, defining it as the area of a sphere of radius R, on each point of which the centre of a solvent molecule can be placed in contact with an atom without penetrating any other atoms of the molecule. The radius R, is given by the sum of the van der Waals radius of the atom and the chosen radius of the solvent molecule (**Figure 3**). The problem with this definition is that it implies that the system is static: it does not account for any movement or flexibility that an atom or group may possess within the molecule.

The deposition of the three-dimensional coordinates of protein structures (solved by X-ray crystallography and nuclear magnetic resonance) in the Protein Data Bank (PDB) Bernstein *et al.*, 1977), has permitted the analysis of relatively large numbers of multimeric proteins. Many
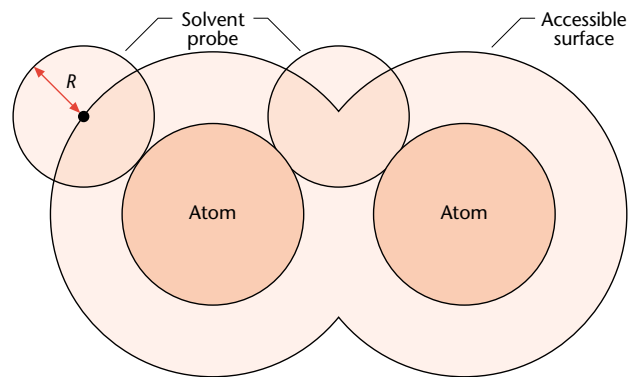

Figure 3 The accessible surface area of a protein. The diagram shows just two atoms, with a probe sphere (with a radius of R) defining the accessible surface.

computational studies (Miller *et al.*, 1987; Argos, 1988; Jones and Thornton, 1996) have analysed the properties of the recognition sites of multimeric proteins in comparison with the protein exterior and protein interior. Such studies have looked for common trends in terms of the size and shape, amino acid composition, hydrogen bonding and secondary structure.

Size and shape

Subunits in protein dimers contribute 6–40% of their ASA to the contact interface; the mean is 12%. For trimers and tetramers the means are 17% and 21%, respectively (Miller *et al.*, 1987). It has been predicted that 5–6% of ASA of the subunit must be contributed to the contact interface as a minimum requirement for its stabilization (Argos, 1988). In terms of absolute ASA buried by each subunit, the ranges for dimers is very large, with small areas recorded for structures such as 434 repressor (368 $Å^2$) and large areas for structures such as in citrate synthase (4746 $Å^2$) (Jones and Thornton, 1995). In homodimers the ASA is approximately linearly related to the molecular weight of the protomer. Thus the larger the protomer, the larger the interface site required to stabilize its interaction with a second protein subunit.

When viewed as an overall or global cross-section, interfaces are generally flat; however, exceptions have been noted. These include proteins in which the two subunits twist together across the interface (e.g. isocitrate dehydrogenase) or proteins that have subunits with 'arms' that clasp the two halves of the structure together (e.g. aspartate aminotransferase).

Amino acid composition

The protein–protein interfaces in multimeric proteins are largely hydrophobic. These interfaces have been shown to be more hydrophobic than the exterior but less hydrophobic than the interior. Calculating residue interface

propensities, which give an indication of the relative importance of different amino acids in the interface compared with the protein surface as a whole, reveals that specific amino acids have high probabilities of being present in protein–protein interfaces compared with their frequency on the exposed surface of the protein (Jones and Thornton, 1996). The hydrophobic residues occur frequently in the interfaces, along with the single aromatic residues, histidine, tyrosine and phenylalanine, which make particularly good 'glue' for sticking together protein subunits (Argos, 1988).

Electrostatic interactions

The number of intermolecular hydrogen bonds is approximately proportional to the ASA buried in the interface. In homodimers there are, on average, 0.88 hydrogen bonds per $100\,\text{Å}^2$ of ASA buried (for interfaces covering $> 1500\,\text{Å}^2$ per subunit); but the number of hydrogen bonds varies from zero in some complexes (e.g. uteroglobin) to as many as 46 in variant surface glycoprotein (Jones and Thornton, 1995). Side-chain hydrogen bonds represent approximately 76–78% of the interactions. Salt bridges have also been observed between subunits of multimeric proteins, but only 56% of homodimeric proteins were found to possess such interactions, many having none or, at the most, five. Intermolecular disulfides are rarely seen in dimeric proteins, as they only occur in oxidizing environments; however, when intermolecular disulfides do occur they often play an important role in structural stabilization. Protein engineering experiments on two structures, platelet-derived growth factor B and thymidylate synthase, have shown in both that the introduction of intermolecular disulfides increases the stability of the protein associations.

Secondary structure

Interfaces in multimeric proteins occur between helix, sheet and coil motifs, with both like and nonlike interactions observed across the interface. Interfaces commonly have a central area of extended sheet, helix–helix packing or sheet–sheet packing decorated at the edges by loop interactions. The loop interactions contributed on average 40% of the interface contacts (Miller, 1989). The loops commonly interact with other loops and with the ends of secondary structures, and are stabilized by large numbers of hydrogen bonds. Motifs are often shared across interfaces; stability within interfaces is enhanced by converting loops within motifs into linkers across interfaces.

Concluding Remarks

Quaternary structure is the highest level of protein organization. The quaternary structure of a protein is fundamentally important to its functional role in the cell. Chemical and physical properties, including hydrophobic interactions, electrostatic interactions and shape complementarity, play complex roles in the interaction of one protein subunit with another. The importance of these properties varies depending upon the type of complex and its function.

References

Argos P (1988) An investigation of protein subunit and domain interfaces. *Protein Engineering* **2**: 101–113.
Bernstein FC, Koetzle TF, Williams GJB *et al.* (1977) The protein data bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **112**: 535–542.
Chothia C and Janin J (1975) Principles of protein–protein recognition. *Nature* **256**: 705–708.
Fersht AR (1987) The hydrogen bond in molecular recognition. *Trends in Biochemical Sciences* **12**: 301–304.
Jaenicke R (1987) Folding and association of proteins. *Progress in Biophysics and Molecular Biology* **29**: 117–237.
Jones S and Thornton JM (1995) Protein–protein interactions: a review of protein dimer structures. *Progress in Biophysics and Molecular Biology* **63**: 31–65.
Jones S and Thornton JM (1996) Principles of protein–protein interactions. *Proceedings of the National Academy of Sciences of the USA* **93**: 13–20.
Lapthorn AJ, Harris DC, Littlejohn A *et al.* (1994) Crystal structure of human chorionic gonadotropin. *Nature* **369**: 455–461.
Lawrence MC and Colman PM (1993) Shape complementarity at protein/protein interfaces. *Journal of Molecular Biology* **234**: 946–950.
Lee B and Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55**: 379–400.
Miller S (1989) The structure of interfaces between subunits of dimeric and tetrameric proteins. *Protein Engineering* **3**: 77–83.
Miller S, Lesk AM, Janin J and Chothia C (1987) The accessible surface area and stability of oligomeric proteins. *Nature* **328**: 834–836.
Monod J, Wyman J and Changeux J (1965) On the nature of allosteric transitions: a plausible model. *Journal of Molecular Biology* **12**: 88–118.
Navia MA, Fitzgerald PMD, McKeever BM *et al.* (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**: 615–620.

Further Reading

Banaszak LJ, Birktoft JJ and Barry CD (1981) Protein–protein interactions and protein structures. In: *Protein–Protein Interactions*, pp. 31–128. New York: Wiley.
Chan WW (1976) The relationship between quaternary structure and enzyme activity. *Trends in Biological Sciences* **11**: 258–260.
Duquerroy S, Cherfils J and Janin J (1991) Protein–protein interaction: an analysis by computer simulation. In: Chadwick DJ and Widdows K (eds) *Protein Conformation*, pp. 237–252. Chichester, UK: Wiley.
Garel JR (1992) Folding of large proteins: multidomain and multisubunit proteins. In: Creighton TE (ed.) *Protein Folding*, pp. 405–454. New York: Freeman.
Garel JR, Martel A, Muller K *et al.* (1984) Role of subunit interactions in the self-assembly of oligomeric proteins. *Advances in Biophysics* **18**: 91–113.

Klotz IM, Darnell DW and Langerman NR (1975) Quaternary structure of proteins. In: Neurath H and Hill RL (eds) *The Proteins*, 3rd edn, pp. 25–62. New York: Academic Press.

Matthews BW and Bernhard SA (1973) Structure and symmetry of oligomeric enzymes. *Annual Review of Biophysics and Bioengineering* **2**: 257–317.

Riddihough G (1994) The evolution of oligomerization. *Nature Structural Biology* **1**: 411–412.

Weber G (1992) *Protein Interactions*. London: Chapman and Hall.