1

An Introduction to protein structure and function

Biochemistry has exploded as a major scientific endeavour over the last one hundred years to rival previously established disciplines such as chemistry and physics. This occurred with the recognition that living systems are based on the familiar elements of organic chemistry (carbon, oxygen, nitrogen and hydrogen) together with the occasional involvement of inorganic chemistry and elements such as iron, copper, sodium, potassium and magnesium. More importantly the laws of physics including those concerning thermodynamics, electricity and quantum physics are applicable to biochemical systems and no 'vital' force distinguishes living from non-living systems. As a result the laws of chemistry and physics are successfully applied to biochemistry and ideas from physics and chemistry have found widespread application, frequently revolutionizing our understanding of complex systems such as cells.

This book focuses on one major component of all living systems – the proteins. Proteins are found in all living systems ranging from bacteria and viruses through the unicellular and simple eukaryotes to vertebrates and higher mammals such as humans. Proteins make up over 50 percent of the dry weight of cells and are present in greater amounts than any other biomolecule. Proteins are unique amongst the macromolecules in underpinning every reaction occurring in biological systems. It goes without saying that one should not ignore the other components of living systems since they have indispensable roles, but in this text we will consider only proteins.

A brief and very selective historical perspective

With the vast accumulation of knowledge about proteins over the last 50 years it is perhaps surprising to discover that the term *protein* was introduced nearly 170 years ago. One early description was by Gerhardus Johannes Mulder in 1839 where his studies on the composition of animal substances, chiefly fibrin, albumin and gelatin, showed the presence of carbon, hydrogen, oxygen and nitrogen. In addition he recognized that sulfur and phosphorus were present sometimes in 'animal substances' that contained large numbers of atoms. In other words, he established that these 'substances' were macromolecules. Mulder communicated his results to Jöns Jakob Berzelius and it is suggested the term protein arose from this interaction where the origin of the word protein has been variously ascribed to derivation from the Latin word *primarius* or from the Greek god *Proteus*. The definition of proteins was timely since in 1828 Friedrich Wöhler had shown that

Proteins: Structure and Function by David Whitford

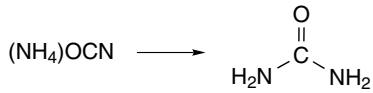

Figure 1.1 The decomposition of ammonium cyanate yields urea

heating ammonium cyanate resulted in isomerism and the formation of urea (Figure 1.1). Organic compounds characteristic of living systems, such as urea, could be derived from simple inorganic chemicals. For many historians this marks the beginning of biochemistry and it is appropriate that the discovery of proteins occurred at the same period.

The development of biochemistry and the study of proteins was assisted by analysis of their composition and structure by Heinrich Hlasiwetz and Josef Habermann around 1873 and the recognition that proteins were made up of smaller units called amino acids. They established that hydrolysis of casein with strong acids or alkali yielded glutamic acid, aspartic acid, leucine, tyrosine and ammonia whilst the hydrolysis of other proteins yielded a different group of products. Importantly their work suggested that the properties of proteins depended uniquely on the constituent parts – a theme that is equally relevant today in modern biochemical study.

Another landmark in the study of proteins occurred in 1902 with Franz Hofmeister establishing the constituent atoms of the peptide bond with the polypeptide backbone derived from the condensation of free amino acids. Five years earlier Eduard Buchner revolutionized views of protein function by demonstrating that yeast cell *extracts* catalysed fermentation of sugar into ethanol and carbon dioxide. Previously it was believed that only living systems performed this catalytic function. Emil Fischer further studied biological catalysis and proposed that components of yeast, which he called enzymes, combined with sugar to produce an intermediate compound. With the realization that cells were full of enzymes 100 years of research has developed and refined these discoveries. Further landmarks in the study of proteins could include Sumner's crystallization of the first enzyme (urease) in 1926 and Pauling's description of the geometry of the peptide bond; however, extensive discussion of these advances and many other important discoveries in protein biochemistry are best left to history of science textbooks.

A brief look at the award of the Nobel Prizes for Chemistry, Physiology and Medicine since 1900 highlighted in Table 1.1 reveals the involvement of many diverse areas of science in protein biochemistry. At first glance it is not obvious why William and Lawrence Bragg's discovery of the diffraction of X-rays by sodium chloride crystals is relevant, but diffraction by protein crystals is the main route towards biological structure determination. Their discovery was the first step in the development of this technique. Discoveries in chemistry and physics have been implemented rapidly in the study of proteins. By 1958 Max Perutz and John Kendrew had determined the first protein structure and this was soon followed by the larger, multiple subunit, structure of haemoglobin and the first enzyme, lysozyme. This remarkable advance in knowledge extended from initial understanding of the atomic composition of proteins around 1900 to the determination of the three-dimensional structure of proteins in the 1960s and represents a major chapter of modern biochemistry. However, advances have continued with new areas of molecular biology proving equally important to understanding protein structure and function.

Life may be defined as the ordered interaction of proteins and all forms of life from viruses to complex, specialized, mammalian cells are based on proteins made up of the same building blocks or amino acids. Proteins found in simple unicellular organisms such as bacteria are identical in structure and function to those found in human cells illustrating the evolutionary lineage from simple to complex organisms.

Molecular biology starts with the dramatic elucidation of the structure of the DNA double helix by James Watson, Francis Crick, Rosalind Franklin and Maurice Wilkins in 1953. Today, details of DNA replication, transcription into RNA and the synthesis of proteins (translation) are extensive. This has established an enormous body of knowledge representing a whole new subject area. All cells encode the information content of proteins within genes, or more accurately the order of bases along the DNA strand, yet it is the

Table 1.1 Selected landmarks in the study of protein structure and function from 1900–2002 as seen by the award of the Nobel Prize for Chemistry, Physiology or Medicine

Date	Discoverer + Discovery
1901	Wilhelm Conrad Röntgen 'in recognition of the ... discovery of the remarkable rays subsequently named after him'
1907	Eduard Buchner 'cell-free fermentation'
1914	Max von Laue 'for his discovery of the diffraction of X-rays by crystals'
1915	William Henry Bragg and William Lawrence Bragg 'for their services in the analysis of crystal structure by ... X-rays'
1923	Frederick Grant Banting and John James Richard Macleod 'for the discovery of insulin'
1930	Karl Landsteiner 'for his discovery of human blood groups'
1946	James Batcheller Sumner 'for his discovery that enzymes can be crystallized'.
	John Howard Northrop and Wendell Meredith Stanley 'for their preparation of enzymes and virus proteins in a pure form'
1948	Arne Wilhelm Kaurin Tiselius 'for his research on electrophoresis and adsorption analysis, especially for his discoveries concerning the complex nature of the serum proteins'
1952	Archer John Porter Martin and Richard Laurence Millington Synge 'for their invention of partition chromatography'
1952	Felix Bloch and Edward Mills Purcell 'for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith'
1954	Linus Carl Pauling 'for his research into the nature of the chemical bond and ... to the elucidation of ... complex substances'
1958	Frederick Sanger 'for his work on the structure of proteins, especially that of insulin'
1959	Severo Ochoa and Arthur Kornberg 'for their discovery of the mechanisms in the biological synthesis of ribonucleic acid and deoxyribonucleic acid'
1962	Max Ferdinand Perutz and John Cowdery Kendrew 'for their studies of the structures of globular proteins'
1962	Francis Harry Compton Crick, James Dewey Watson and Maurice Hugh Frederick Wilkins 'for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material'
1964	Dorothy Crowfoot Hodgkin 'for her determinations by X-ray techniques of the structures of important biochemical substances'
1965	François Jacob, André Lwoff and Jacques Monod 'for discoveries concerning genetic control of enzyme and virus synthesis'
1968	Robert W. Holley, Har Gobind Khorana and Marshall W. Nirenberg 'for ... the genetic code and its function in protein synthesis'
1969	Max Delbrück, Alfred D. Hershey and Salvador E. Luria 'for their discoveries concerning the replication mechanism and the genetic structure of viruses'

(*continued overleaf*)

Table 1.1 (*continued*)

Date	Discoverer + Discovery
1972	Christian B. Anfinsen 'for his work on ribonuclease, especially concerning the connection between the amino acid sequence and the biologically active conformation' Stanford Moore and William H. Stein 'for their contribution to the understanding of the connection between chemical structure and catalytic activity of . . . ribonuclease molecule'
1972	Gerald M. Edelman and Rodney R. Porter 'for their discoveries concerning the chemical structure of antibodies'
1975	John Warcup Cornforth 'for his work on the stereochemistry of enzyme-catalyzed reactions'. Vladimir Prelog 'for his research into the stereochemistry of organic molecules and reactions'
1975	David Baltimore, Renato Dulbecco and Howard Martin Temin 'for their discoveries concerning the interaction between tumour viruses and the genetic material of the cell'
1978	Werner Arber, Daniel Nathans and Hamilton O. Smith 'for the discovery of restriction enzymes and their application to problems of molecular genetics'
1980	Paul Berg 'for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA' Walter Gilbert and Frederick Sanger 'for their contributions concerning the determination of base sequences in nucleic acids'
1982	Aaron Klug 'development of crystallographic electron microscopy and structural elucidation of nucleic acid–protein complexes'
1984	Robert Bruce Merrifield 'for his development of methodology for chemical synthesis on a solid matrix'
1984	Niels K. Jerne, Georges J.F. Köhler and César Milstein 'for theories concerning the specificity in development and control of the immune system and the discovery of the principle for production of monoclonal antibodies'
1988	Johann Deisenhofer, Robert Huber and Hartmut Michel 'for the determination of the structure of a photosynthetic reaction centre'
1989	J. Michael Bishop and Harold E. Varmus 'for their discovery of the cellular origin of retroviral oncogenes'
1991	Richard R. Ernst 'for . . . the methodology of high resolution nuclear magnetic resonance spectroscopy'
1992	Edmond H. Fischer and Edwin G. Krebs 'for their discoveries concerning reversible protein phosphorylation as a biological regulatory mechanism'
1993	Kary B. Mullis 'for his invention of the polymerase chain reaction (PCR) method' and Michael Smith 'for his fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis'
1994	Alfred G. Gilman and Martin Rodbell 'for their discovery of G-proteins and the role of these proteins in signal transduction'

Table 1.1 (*continued*)

Date	Discoverer + Discovery
1997	Paul D. Boyer and John E. Walker 'for their elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP)'. Jens C. Skou 'for the first discovery of an ion-transporting enzyme, Na^+, K^+-ATPase'
1997	Stanley B. Prusiner 'for his discovery of prions – a new biological principle of infection'
1999	Günter Blobel 'for the discovery that proteins have intrinsic signals that govern their transport and localization in the cell'
2000	Arvid Carlsson, Paul Greengard and Eric R Kandel 'signal transduction in the nervous system'
2001	Paul Nurse, Tim Hunt and Leland Hartwill 'for discoveries of key regulators of the cell cycle'
2002	Kurt Wuthrich, 'for development of NMR spectroscopy as a method of determining biological macromolecules structure in solution.' John B. Fenn and Koichi Tanaka 'for their development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecules'. Sydney Brenner, H. Robert Horvitz and John E. Sulston 'for their discoveries concerning genetic regulation of organ development and programmed cell death'

conversion of this information or expression into proteins that represents the tangible evidence of a living system or life.

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{protein}$$

Cells divide, synthesize new products, secrete unwanted products, generate chemical energy to sustain these processes via specific chemical reactions, and in all of these examples the common theme is the mediation of proteins.

In 1944 the physicist Erwin Schrödinger posed the question 'What is Life?' in an attempt to understand the physical properties of a living cell. Schrödinger suggested that living systems obeyed all laws of physics and should not be viewed as exceptional but instead reflected the statistical nature of these laws. More importantly, living systems are amenable to study using many of the techniques familiar to chemistry and physics. The last 50 years of biochemistry have demonstrated this hypothesis emphatically with tools developed by physicists and chemists rapidly employed in biological studies. A casual perusal of Table 1.1 shows how quickly methodologies progress from discovery to application.

The biological diversity of proteins

Proteins have diverse biological functions ranging from DNA replication, forming cytoskeletal structures, transporting oxygen around the bodies of multicellular organisms to converting one molecule into another. The types of functional properties are almost endless and are continually being increased as we learn more about proteins. Some important biological functions are outlined in Table 1.2 but it is to be expected that this rudimentary list of properties will expand each year as new proteins are characterized. A formal demarcation of proteins into one class should not be pursued too far since proteins can have multiple roles or functions; many proteins do not lend themselves easily to classification schemes. However, for all chemical reactions occurring in cells a protein is involved intimately in the biological process. These proteins are united through their composition based on the same group of 20 amino acids. Although all proteins are composed of the same group of 20 amino acids they differ in their composition – some contain a surfeit of one amino acid whilst others may lack one or two members of the group of 20 entirely. It was realized early in the study of proteins that

Table 1.2 A selective list of some functional roles for proteins within cells

Function	Examples
Enzymes or catalytic proteins	Trypsin, DNA polymerases and ligases,
Contractile proteins	Actin, myosin, tubulin, dynein,
Structural or cytoskeletal proteins	Tropocollagen, keratin,
Transport proteins	Haemoglobin, myoglobin, serum albumin, ceruloplasmin, transthyretin
Effector proteins	Insulin, epidermal growth factor, thyroid stimulating hormone,
Defence proteins	Ricin, immunoglobulins, venoms and toxins, thrombin,
Electron transfer proteins	Cytochrome oxidase, bacterial photosynthetic reaction centre, plastocyanin, ferredoxin
Receptors	CD4, acetycholine receptor,
Repressor proteins	Jun, Fos, Cro,
Chaperones (accessory folding proteins)	GroEL, DnaK
Storage proteins	Ferritin, gliadin,

variation in size and complexity is common and the molecular weight and number of subunits (polypeptide chains) show tremendous diversity. There is no correlation between size and number of polypeptide chains. For example, insulin has a relative molecular mass of 5700 and contains two polypeptide chains, haemoglobin has a mass of approximately 65 000 and contains four polypeptide chains, and hexokinase is a single polypeptide chain with an overall mass of ~100 000 (see Table 1.3).

The molecular weight is more properly referred to as the relative molecular mass (symbol M_r). This is defined as the mass of a molecule relative to 1/12th the mass of the carbon (^{12}C) isotope. The mass of this isotope is defined as exactly 12 atomic mass units. Consequently the term molecular weight or relative molecular mass is a dimensionless quantity and should not possess any units. Frequently in this and many other textbooks the unit Dalton (equivalent to 1 atomic mass unit, i.e. 1 Dalton = 1 amu) is used and proteins are described with molecular weights of 5.5 kDa (5500 Daltons). More accurately, this is the absolute molecular weight representing the mass in grams of 1 mole of protein. For most purposes this becomes of little relevance and the term 'molecular weight' is used freely in protein biochemistry and in this book.

Table 1.3 The molecular masses of proteins together with the number of subunits. The term 'subunit' is synonymous with the number of polypeptide chains and is used interchangeably

Protein	Molecular mass	Subunits
Insulin	5700	2
Haemoglobin	64 500	4
Tropocollagen	285 000	3
Subtilisin	27 500	1
Ribonuclease	12 600	1
Aspartate transcarbamoylase	310 000	12
Bacteriorhodopsin	26 800	1
Hexokinase	102 000	1

Proteins are joined covalently and non-covalently with other biomolecules including lipids, carbohydrates,

nucleic acids, phosphate groups, flavins, heme groups and metal ions. Components such as hemes or metal ions are often called prosthetic groups. Complexes formed between lipids and proteins are lipoproteins, those with carbohydrates are called glycoproteins, whilst complexes with metal ions lead to metalloproteins, and so on. The complexes formed between metal ions and proteins increases the involvement of elements of the periodic table beyond that expected of typical organic molecules (namely carbon, hydrogen, nitrogen and oxygen). Inspection of the periodic table (Figure 1.2) shows that at least 20 elements have been implicated directly in the structure and function of proteins (Table 1.4). Surprisingly elements such as aluminium and silicon that are very abundant in the Earth's crust (8.1 and 25.7 percent by weight, respectively) do not occur in high concentration within cells. Aluminium is rarely, if ever, found as part of proteins whilst the role of silicon is confined to biomineralization where it is the core component of shells. The involvement of carbon, hydrogen, oxygen, nitrogen, phosphorus and sulfur is clear although the role of other elements, particularly transition metals, has been difficult to establish. Where transition metals occur in proteins there is frequently only one metal atom per mole of protein and led in the past to a failure to detect metal. Other elements have an inferred involvement from growth studies showing that depletion from the diet leads to an inhibition of normal cellular function. For metalloproteins the absence of the metal can lead to a loss of structure and function.

Metals such as Mo, Co and Fe are often found associated with organic co-factors such as pterin, flavins, cobalamin and porphyrin (Figure 1.3). These organic ligands hold metal centres and are often tightly associated to proteins.

Table 1.4 The involvement of trace elements in the structure and function of proteins

Element	Functional role
Sodium	Principal intracellular ion, osmotic balance
Potassium	Principal intracellular ion, osmotic balance
Magnesium	Bound to ATP/GTP in nucleotide binding proteins, found as structural component of hydrolase and isomerase enzymes
Calcium	Activator of calcium binding proteins such as calmodulin
Vanadium	Bound to enzymes such as chloroperoxidase.
Manganese	Bound to pterin co-factor in enzymes such as xanthine oxidase or sulphite oxidase. Also found in nitrogenase and as component of water splitting enzyme in higher plants.
Iron	Important catalytic component of heme enzymes involved in oxygen transport as well as electron transfer. Important examples are haemoglobin, cytochrome oxidase and catalase.
Cobalt	Metal component of vitamin B_{12} found in many enzymes.
Nickel	Co-factor found in hydrogenase enzymes
Copper	Involved as co-factor in oxygen transport systems and electron transfer proteins such as haemocyanin and plastocyanin.
Zinc	Catalytic component of enzymes such as carbonic anhydrase and superoxide dismutase.
Chlorine	Principal intracellular anion, osmotic balance
Iodine	Iodinated tyrosine residues form part of hormone thyroxine and bound to proteins
Selenium	Bound at active centre of glutathione peroxidase

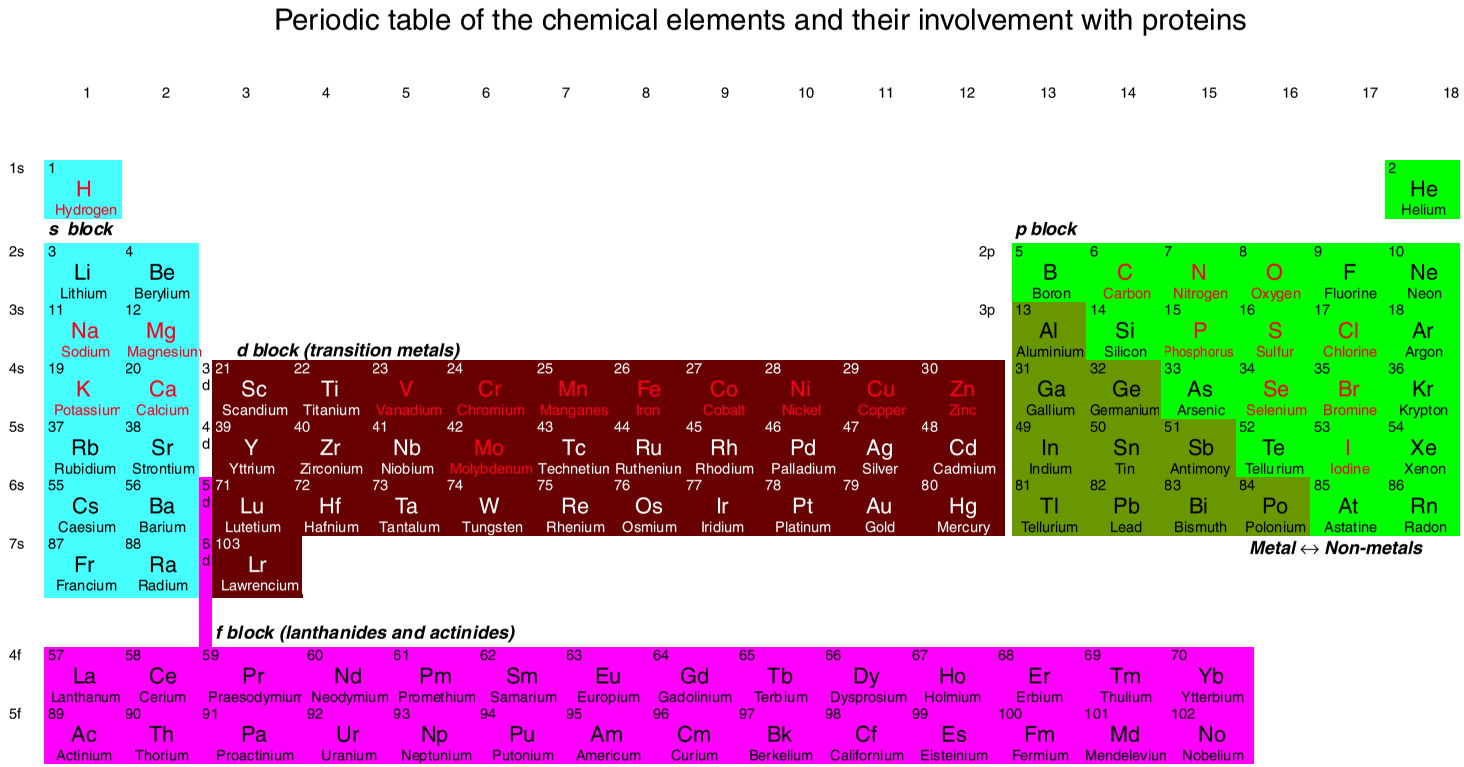

Figure 1.2 The periodic table showing the elements highlighted in red known to have involvement in the structure and/or function of proteins. The involvement of some elements is contentious tungsten and cadmium are claimed to be associated with proteins yet these elements are also known to be toxic

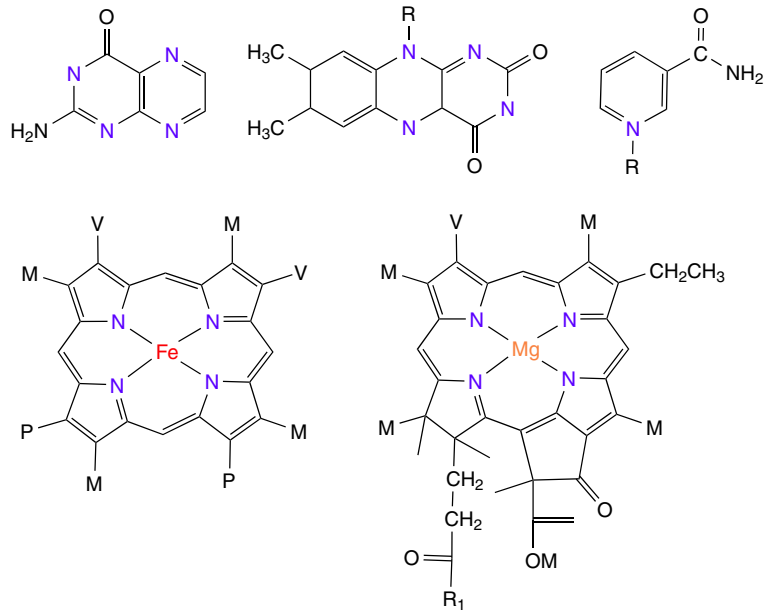

Figure 1.3 Organic co-factors found in proteins. These co-factors are pterin, the isoalloxine ring found as part of flavin in FAD and FMN, the pyridine ring of NAD and its close analogue NADP and the porphyrin skeletons of heme and chlorophyll. R represents the remaining part of the co-factor whilst M and V signify methyl and vinyl side chains

Proteins and the sequencing of the human and other genomes

Recognition of the diverse roles of proteins in biological systems increased largely as a result of the enormous amount of sequencing information generated via the Human Genome Mapping project. Similar schemes aimed at deciphering the genomes of *Escherichia coli*, yeast (*Sacharromyces cerevisiae*), and mouse provided related information. With the completion of the first draft of the human genome mapping project in 2001 human chromosomes contain approximately 25–30 000 genes. This allows a conservative estimate of the number of polypeptides making up most human cells as ~25 000, although alternative splicing of genes and variations in subunit composition increase the number of proteins further. Despite sequencing the human genome it is an unfortunate fact that we do not know the role performed by most proteins. Of those thousands of polypeptides we know the structures of only a small number, emphasizing a large imbalance between the abundance of sequence data and the presence of structure/function information. An analysis of protein databases suggests about 1000 distinct structures or folds have been determined for globular proteins. Many proteins are retained within cell membranes and we know virtually nothing about the structures of these proteins and only slightly more about their functional roles. This observation has enormous consequences for understanding protein structure and function.

Why study proteins?

This question is often asked not entirely without reason by many undergraduates during their first introduction to the subject. Perhaps the best reply that can be given is that proteins underpin every aspect of biological activity. This is particularly important in areas where protein structure and function have an impact on human endeavour such as medicine. Advances in molecular genetics reveal that many diseases stem from specific protein defects. A classic example is cystic

Figure 1.4 The shape of erythrocytes in normal and sickle cell anemia arises from mutations to haemoglobin found within the red blood cell. (Reproduced with permission from Voet, D, Voet, J.G and Pratt, C.W. *Fundamentals of Biochemistry*. John Wiley & Sons Inc.)

fibrosis, an inherited condition that alters a protein, called the cystic fibrosis transmembrane conductance regulator (CFTR), involved in the transport of sodium and chloride across epithelial cell membranes. This defect is found in Caucasian populations at a ratio of ~1 in 20, a surprisingly high frequency. With 1 in 20 of the population 'carrying' a single defective copy of the gene individuals who inherit defective copies of the gene from each parent suffer from the disease. In the UK the incidence of cystic fibrosis is approximately 1 in 2000 live births, making it one of the most common inherited disorders. The disease results in the body producing a thick, sticky mucus that blocks the lungs, leading to serious infection, and inhibits the pancreas, stopping digestive enzymes from reaching the intestines where they are required to digest food. The severity of cystic fibrosis is related to *CFTR* gene mutation, and the most common mutation, found in approximately 65 percent of all cases, involves the deletion of a single amino acid residue from the protein at position 508. A loss of one residue out of a total of nearly 1500 amino acid residues results in a severe decrease in the quality of life with individuals suffering from this disease requiring constant medical care and supervision.

Further examples emphasize the need to understand more about proteins. The pioneering studies of Vernon Ingram in the 1950s showed that sickle cell anemia arose from a mutation in the β chain of haemoglobin. Haemoglobin is a tetrameric protein containing 2α and 2β chains. In each of the β chains a mutation is found that involves the change of the sixth amino acid residue from a glutamic acid to a valine. The alteration of two residues out of 574 leads to a drastic change in the appearance of red blood cells from their normal biconcave disks to an elongated sickle shape (Figure 1.4).

As the name of the disease suggests individuals are anaemic showing decreased haemoglobin content in red blood cells from approximately 15 g per 100 ml to under half that figure, and show frequent illness. Our understanding of cystic fibrosis and of sickle cell anaemia has advanced in parallel with our understanding of protein structure and function although at best we have very limited and crude means of treating these diseases.

However, perhaps the greatest impetus to understand protein structure and function lies in the hope of overcoming two major health issues confronting the world in the 21st century. The first of these is cancer. Cancer is the uncontrolled proliferation of cells that have lost their normal regulated cell division often in response to a genetic or environmental trigger. The development of cancer is a multistep, multifactorial process often occurring over decades but the precise involvement of specific proteins has been demonstrated in some instances. One of the best examples is a protein called p53, normally present at low levels in cells, that 'switches on' in response to cellular damage and as a transcription factor controls the cell cycle process. Mutations in p53 alter the normal cycle of events leading eventually to cancer and several tumours

including lung, colorectal and skin carcinomas are attributed to molecular defects in p53. Future research on p53 will enable its physicochemical properties to be thoroughly appreciated and by understanding the link between structure, folding, function and regulation comes the prospect of unravelling its role in tumour formation and manipulating its activity via therapeutic intervention. Already some success is being achieved in this area and the future holds great promise for 'halting' cancer by controlling the properties of p53 and similar proteins.

A second major problem facing the world today is the estimated number of people infected with the human immunodeficiency virus (HIV). In 2003 the World Health Organization (WHO) estimated that over 40 million individuals are infected with this virus in the world today. For many individuals, particularly those in the 'Third World', the prospect of prolonged good health is unlikely as the virus slowly degrades the body's ability to fight infection through damage to the immune response mechanism and in particular to a group of cells called cytotoxic T cells. HIV infection encompasses many aspects of protein structure and function, as the virus enters cells through the interaction of specific viral coat proteins with receptors on the surface of white blood cells. Once inside cells the virus 'hides' but is secretly replicating and integrating genetic material into host DNA through the action of specific enzymes (proteins). Halting the destructive influence of HIV relies on understanding many different, yet inter-related, aspects of protein structure and function. Again, considerable progress has been made since the 1980s when the causative agent of the disease was recognized as a retrovirus. These advances have focussed on understanding the structure of HIV proteins and in designing specific inhibitors of, for example, the reverse transcriptase enzyme. Although in advanced health care systems these drugs (inhibitors) prolong life expectancy, the eradication of HIV's destructive action within the body and hence an effective cure remains unachieved. Achieving this goal should act as a timely reminder for all students of biology, chemistry and medicine that success in this field will have a dramatic impact on the quality of human life in the forthcoming decades.

Central to success in treating any of the above diseases are the development of new medicines, many based on proteins. The development of new therapies has been rapid during the last 20 years with the list of new treatments steadily increasing and including minimizing serious effects of different forms of cancer via the use of specific proteins including monoclonal antibodies, alleviating problems associated with diabetes by the development of improved recombinant 'insulins' and developing 'clot-busting' drugs (proteins) for the management of strokes and heart attacks. This highly selective list is the productive result of understanding protein structure and function and has contributed to a marked improvement in disease management. For the future these advances will need to be extended to other diseases and will rely on an extensive and thorough knowledge of proteins of increasing size and complexity. We will need to understand the structure of proteins, their interaction with other biomolecules, their roles within different biological systems and their potential manipulation by genetic or chemical methods. The remaining chapters in this book represent an attempt to introduce and address some of these issues in a fundamental manner helpful to students.