

## Chapter 2

# Proteins, Enzymes and Biological Catalysis

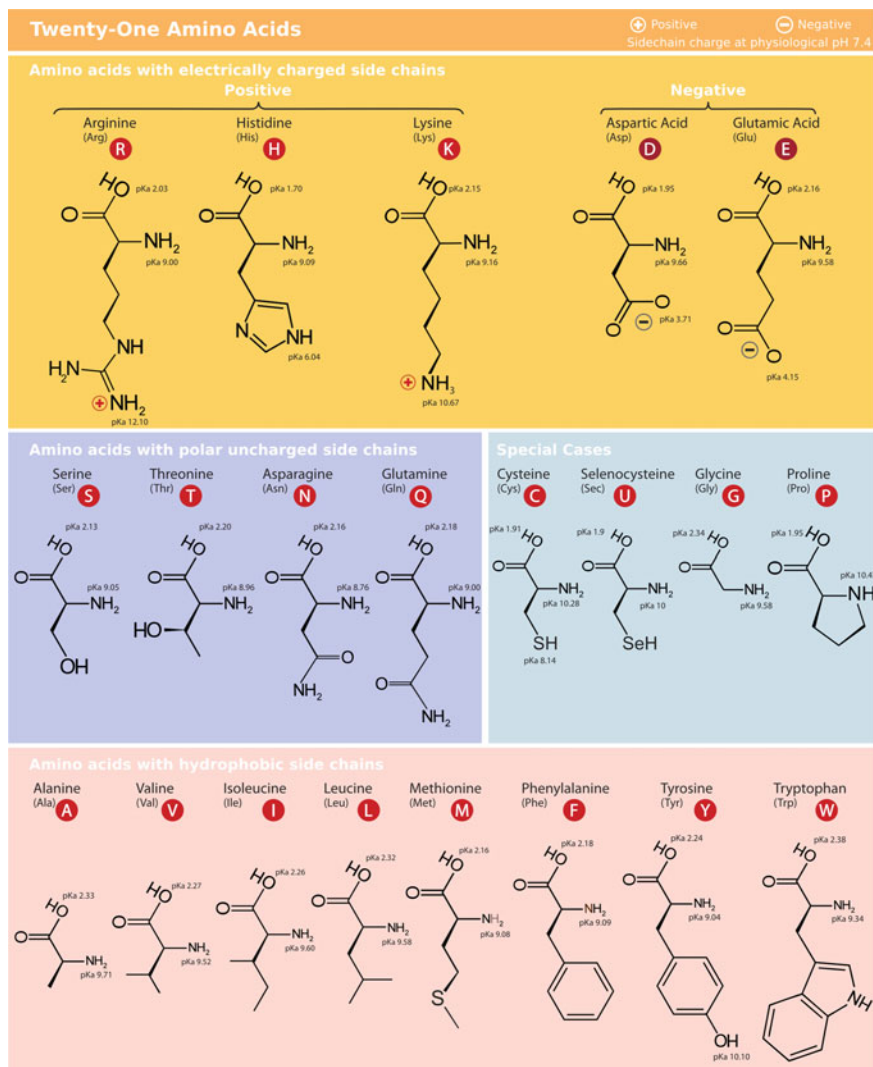
*In biology, proteins are uniquely important... the most significant thing about proteins is that they can do almost anything. But their main function is to act as enzymes*

—Francis Crick, *Society for Experimental Biology Symposium* 1957

In their simplest form, proteins essentially comprise unbranched polymer chains formed as a result of chemical bonding that takes place between the amino acid building blocks. This sequence of building blocks can be readily and rapidly determined via experimental means, building upon the pioneering work of Frederick Sanger who, in 1951, obtained the amino acid sequence of insulin [1], the first protein to have its sequence determined. However, these sequences give as much information about the biology of the system as a London telephone directory gives about the function and wonder of the city. The primary sequence formed by the amino acids then forms complex secondary structure through an intricate process of folding. The resulting structure of the protein in turn defines its function and in the case of enzymes, the particular type of reaction it catalyses. The objective of this chapter is to outline the key biological concepts used throughout this dissertation.

## 2.1 Amino Acids

Hydrogen, carbon, nitrogen and oxygen constitute 96.5 % of the mass of living cells [2]. This rises to 98 % when taking sulphur and phosphorous into consideration. Therefore, it is clear that the chemistry of life is dominated by the lighter elements. It is these elements that also form the amino acids, or residues, which are the subunits of proteins and can be seen in Fig. 2.1. The general chemical formula for an amino acid is  $\text{NH}_2\text{C}^\alpha\text{RHC}'\text{O}_2\text{H}$ . The four-fold coordinated central alpha-carbon atom ( $\text{C}^\alpha$ )



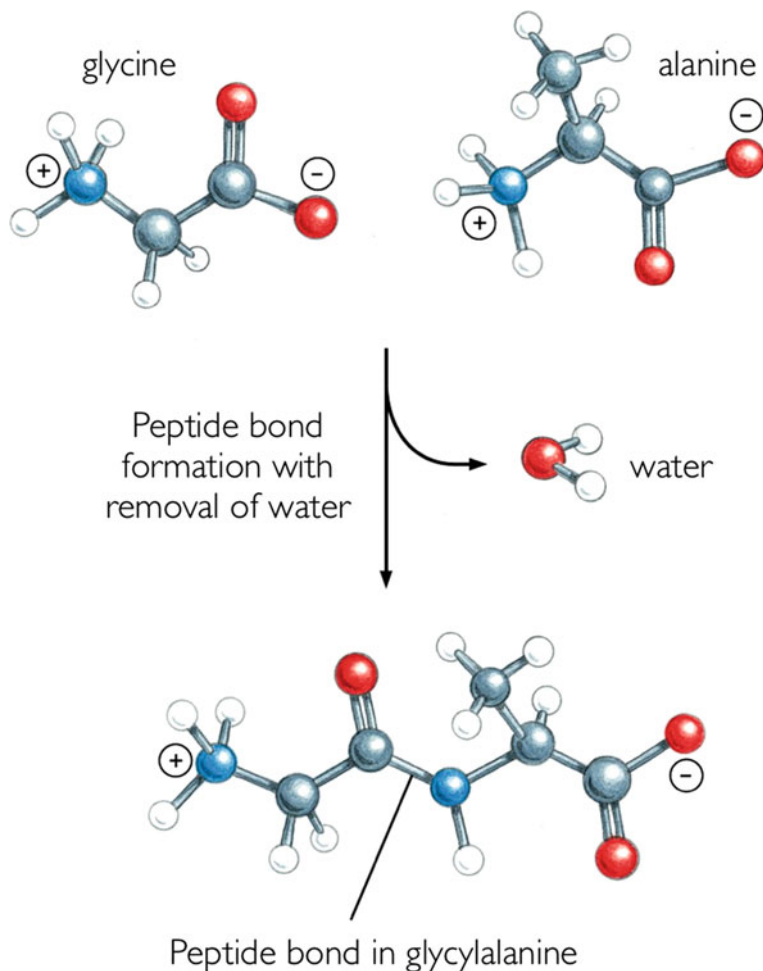
**Fig. 2.1** The twenty-one naturally occurring amino acids. Charged, polar uncharged and hydrophobic side chains are also highlighted. Figure adapted, and pKa data acquired, from Ref. [3]

is  $sp^3$  hybridised and is attached to a hydrogen atom, along with the amino group ( $\text{NH}_2$ ) and carboxylic acid group ( $\text{C}'\text{O}_2\text{H}$ ) and a side chain ( $R$ ) by  $\sigma$  bonds. This bonding pattern is common to all the twenty-one amino acids. What distinguishes one amino acid from another is the side chain ( $R$ ) attached to the alpha-carbon. The amino acids are usually divided into three classes, depending on the chemical nature of the side chain. Classes consist of amino acids with strictly hydrophobic side chains, those with charged residues and those with polar side chains. Most of the twenty-one

naturally occurring amino acids were discovered in the 19th century. How and why exactly this precise set of amino acids came to be chosen as the building blocks of life is one of the mysteries of evolution.

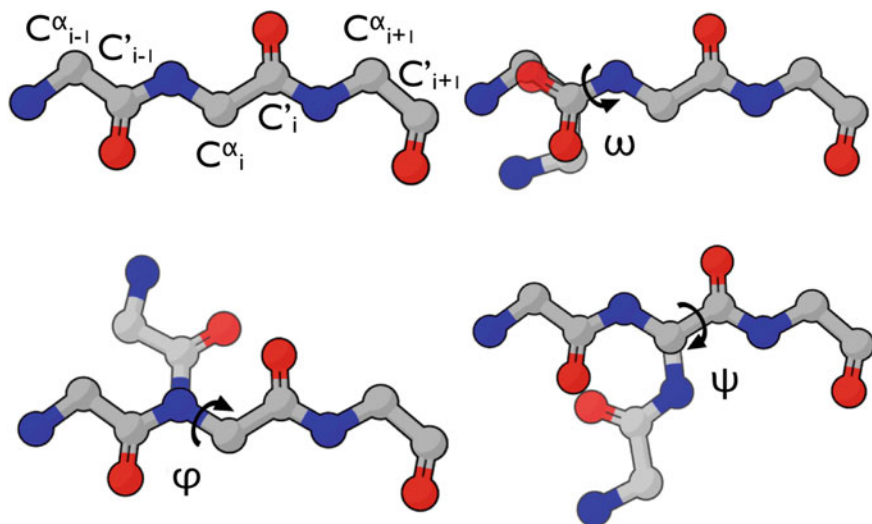
## 2.2 Protein Structure

Proteins are formed in cells and are synthesised in ribosomes. Amino acids are joined into linked chains during this synthesis process when the carboxyl group of one amino acid condenses with the amino group of the next in order to eliminate water. This can be seen in Fig. 2.2. This formation of peptide bonds is repeated as the chain elongates, generating the so-called ‘backbone’ from which the side chains project. The six atoms that surround each peptide bond are constrained in an arrangement close to planar, comprising the alpha carbon ( $C^\alpha$ ), carboxyl carbon ( $C'$ ) and amide nitrogen atoms [4]. The nitrogen, oxygen and subsequent alpha-carbon atoms are also close to coplanar. This is due to the adjacent nitrogen and carbon atoms in the  $N-H-C'=O$  unit being  $sp^2$  hybridised. Their positions and resultant secondary structure can be defined in terms of the angles of rotation about the bonds connecting the three atoms. These angles of rotation are conventionally labeled as  $\psi$ ,  $\phi$  and  $\omega$ , respectively. The peptide backbone dihedral, or torsion, angles are illustrated in Fig. 2.3. The angle  $\phi$  defines the rotation of the plane containing  $C'_i$ ,  $C^\alpha_i$ ,  $O_i$  and  $N_{i+1}$  around the  $N_i-C^\alpha_i$  bond, controlling the  $C'-C'$  distance.  $\psi$  defines the rotation of the plane containing  $C'_i$ ,  $O_i$  and  $N_{i+1}$  around the  $C^\alpha_i-C_i$  bond and controls the  $N-N$  distance.  $\omega$  defines rotation around the peptide bond  $C'-N_{i+1}$  and controls the  $C^\alpha-C^\alpha$  distance but in general is restricted to be close to  $180^\circ$  by the planar nature of the peptide bond, therefore  $\omega$  describes any deviation from planarity. One consequence of the condensation process that leads to the formation of proteins is that the amino group of the first amino acid and the carboxyl group of the last amino acid remain intact. Thus a polypeptide is said to run from its amino ( $N-$ ) terminus to its carboxy ( $C-$ ) terminus. The sequence of amino acids from which a protein is built is termed its primary structure. One of the first important general principles to emerge from protein structure studies was the fact that amino acids in the interior of proteins have almost exclusively hydrophobic side chains. However, in order to form the compact and folded protein structure seen in nature, new interactions are required to compensate for the solvent interactions lost from the peptide background. Thus, there is a major barrier to creating such a hydrophobic core from a protein chain. In order to bring the side chains into the core, the main chain also needs to fold into the interior. Each peptide unit on the backbone has one hydrogen-bond donor (the  $N-H$  group) and one hydrogen-bond acceptor (the  $C'=O$  group) resulting in a very polar and hydrophilic backbone. In order to replace the favourable interactions the backbone would have with the solvent in an unfolded state, a more compact, folded structure is required. Proteins solve this problem by forming secondary structures where the backbone  $N-H$  and  $C'=O$  groups form intramolecular Hydrogen bonds with each other.

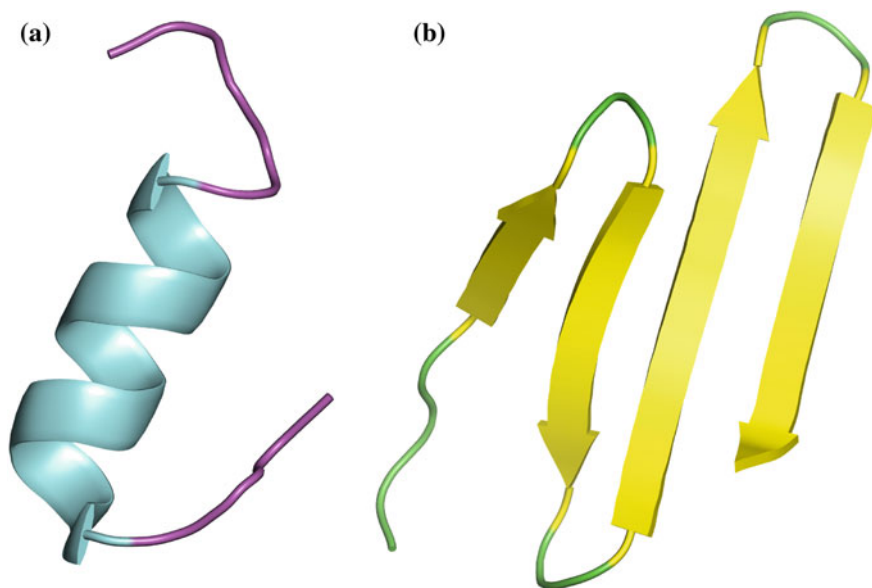


**Fig. 2.2** The condensation process that leads to peptide bond formation. Figure adapted from Ref. [2]

A protein will fold into a stable configuration, or secondary structure, determined by its primary structure of amino acids. Although the secondary structure of proteins can be incredibly varied, there are two commonly recurring motifs. These  $\alpha$  helices and  $\beta$  sheets, as illustrated in Fig. 2.4, are recurring patterns in protein structures and are recognisably similar in virtually all natural proteins, despite varying in size and amino acid composition. These ideas were first put forward by William Astbury in 1933 when investigating keratin and collagen. Astbury proposed that unstretched protein molecules formed a helix, which he called the  $\alpha$ -form, and stretching caused the helix to uncoil, forming an extended state which he called the  $\beta$ -form [5]. Whilst the details of the Astbury model were incorrect, they correspond to the modern ideas



**Fig. 2.3** Peptide backbone dihedral angles

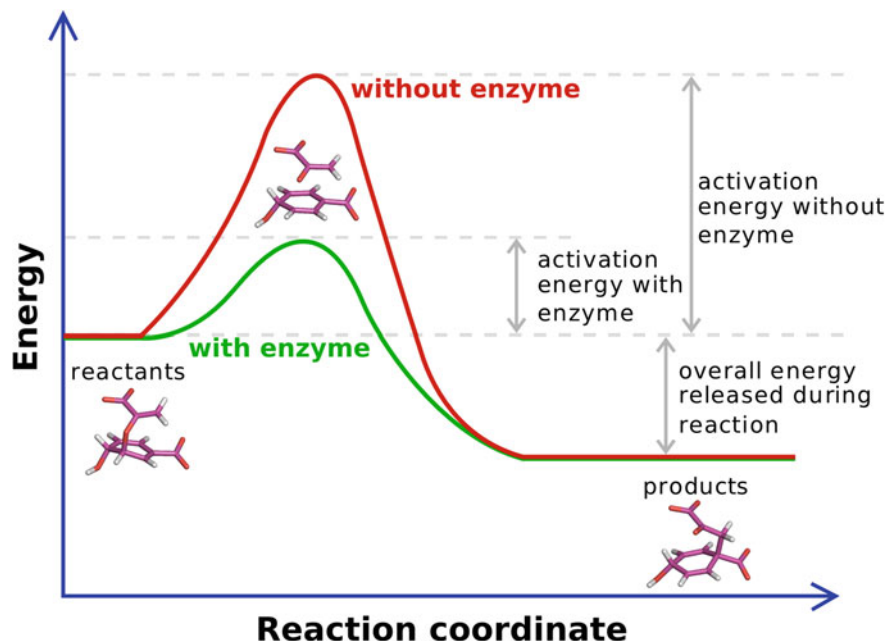


**Fig. 2.4** **a**  $\alpha$  helix and, **b**  $\beta$  sheet secondary structure motifs

of secondary structure which were later refined by Pauling, Robert Corey and Herman Branson in 1951, where Astbury's original  $\alpha$  and  $\beta$  notation was retained [6]. The conformation formed by an entire protein chain, including many secondary structure motifs, is termed its tertiary structure. In addition, if a protein is part of a complex of multiple polypeptide chains then the complete structure is termed the quaternary structure. This last concept remains outside the scope of this dissertation. Proteins are involved in many diverse functions ranging from maintaining the chemical potential across cell membranes to replicating DNA. However, most importantly and most relevant to this dissertation, proteins are actively engaged as enzymes in the catalysis of complex chemical reactions.

## 2.3 Enzyme Catalysis

The cells in a living organism carry out a never-ending series of chemical reactions. This very often involves rearranging small organic molecules in a set of steps along some metabolic pathway. The molecules at the start of this process will usually be the result of photosynthesis in plants or the ingestion of food in mammals. The subsequent pathway will then modify the input molecules sufficiently to meet the requirements of the cells in the living system. Each cell performs many millions of these reactions every second. However, the vast majority of the reactions that take place would normally not happen at the mild temperatures and pressures found in the cell. The key to accelerating, or catalysing, these reactions comes in the form of enzymes. The primary function of an enzyme is to accelerate the reaction rate of a particular chemical reaction relative to the equivalent uncatalysed reaction, or to make a reaction happen that would not occur spontaneously. Enzymes are known to catalyse around 4,000 biochemical reactions [7] with many reaction rates on the order of millions of times faster than the equivalent uncatalysed reactions. The initial ideas laid down by Emil Fischer and his 'lock-and-key' model [8] were used to explain the specificity found in enzymes through the fact that both the enzyme (the 'lock') and the substrate (the 'key') were thought to possess specific complementary geometric shapes that fit exactly into one another. However, while this is an excellent model for describing enzyme-substrate specificity, it does not adequately explain how enzymes manage to catalyse these chemical reactions. The work of Henry Eyring, Meredith Evans and Michael Polanyi [9, 10] in the 1930s, and Linus Pauling [11] in the 1940s, revolutionised the theory of enzyme catalysis by hypothesising transition state structures. It was Pauling's further proposal that the powerful catalytic action of enzymes could be explained by specific tight binding to the transition state species in Ref. [11] that initially led the ideas of transition state stabilisation by enzymes, and would lead the way to the modern transition state theory [12].



**Fig. 2.5** Schematic diagram illustrating how the activation barrier is reduced, compared to the equivalent reaction in the absence of the enzyme, through stabilising the transition state in the presence of the enzyme

It is now widely accepted that enzymes function to stabilise the transition states lying between the reactants and products in the chemical reactions they are catalysing. This stabilisation dramatically reduces the activation energy required for the reaction to take place, therefore greatly accelerating the rate of reaction. An illustration comparing the activation barriers for a reaction in the presence of an enzyme and the equivalent reaction in the absence of the enzyme can be seen in Fig. 2.5. Drawing from the conclusions that an enzyme binds strongly to its particular transition state, the enzyme could also be expected to bind strongly to any synthesised molecule which closely resembles the ionic structure of such a transition state. Whereas reactant and products often participate in several enzyme reactions, the transition state tends to be characteristic of one particular enzyme. Therefore any inhibitor, or transition state analogue, would need to be specific for that particular enzyme. The identification of transition state analogues, for a range of targets [13–16], further supports the transition state stabilisation hypothesis for enzymatic catalysis. How exactly this transition state stabilisation arises is still a topic of debate amongst enzymologists, and the study of the precise mechanisms involved in, and origins of, enzyme catalysis is an active area of research.

Many authors propose that the stabilisation arises mainly due to the favourable Coulombic interactions between the enzyme and the substrate. Therefore, it is crucial to treat the electrostatics of the, often polar, enzyme active site accurately. Enzymes



can alter the electronic structure of their constituent substrates via protonation, proton abstraction, electron transfer, geometric distortion, hydrophobic partitioning and interaction with Lewis acids and bases. This is usually achieved through short-range forces from noncovalent bonds such as van der Waals interactions, electrostatic interactions and hydrogen bonds. A hydrogen bond is the attractive interaction between polar molecules where hydrogen is bound to a highly electronegative atom, such as nitrogen or oxygen, forming an attractive interaction with another atom, such as  $\text{OH} - \cdots \text{N}$ . The hydrogen bond is directional and so is at its strongest when the three atoms involved are aligned. Electrostatic interactions occur between partially charged groups on polar molecules, such as the charged amino acids. At very short distances, any two atoms will show a weak van der Waals interaction, due to their fluctuating electron densities. These three types of weak bonds have less than 1/20 the strength of a standard covalent bond [2]. However, despite a single example of any of these bonds being relatively weak compared to a covalent bond, many of them can form together to create a strong bonding arrangement that stabilises a particular three-dimensional structure. These bonds involve atoms not only in the polypeptide backbone but also the amino acid side chains. The stability of each folded shape is significantly dependent upon the combined strength of large numbers of these noncovalent bonds.

Whilst performing as an enzyme may seem like just another function in the long list of jobs that proteins carry out in the cell, the colossal, unmitigated catalytic power of enzymes is extraordinary [17]. The incredible efficiency demonstrated by the OMPase enzyme, taking a reaction that would otherwise have a half life of 78 million years in solution, to complete in just 18 ms [18] is simply breathtaking. The role of enzymes as biological catalysts is clearly critical for life as just under half of all gene products are annotated as having enzymatic function [19]. In addition, enzymes are often the targets of pharmaceutical development with a significant fraction of approved clinical drugs modifying the behaviour of enzymes implicated in human disease along with disease-causing pathogens [20]. Nearly half of all marketed small molecule therapeutics are designed as enzyme inhibitors [21]. It is argued that ligand design can benefit greatly from improved knowledge of enzyme mechanisms and key active-site interactions [22]. In addition, increasing importance is being given to the prediction of enzyme-mediated adverse reactions [23] and drug metabolism [24]. An understanding of the electronic, atomic and molecular origins of how enzymes achieve their catalytic rate enhancements is a long-standing problem in biochemistry and, increasingly, within computational biology. In many cases, experimental observation alone is not able to establish the mechanisms of enzyme-catalysed reactions and the origins of catalysis, due to a lack of detailed microscopic information regarding the transition state of the reaction in the enzyme.

Transition states are central to many of the fundamental questions that surround chemical reactivity; the stabilisation of such states is a highly important process for the efficiency of catalysis within enzymes. Transition state complexes can often prove very difficult to observe directly in experiment due to their extremely short lifetimes, typically picoseconds. However, it should be noted that the development of femtosecond transition state spectroscopic techniques is currently an active area of



experimental research [25–27]. Computational modelling could, potentially, complement experiment in this task as it has the ability to probe and analyse enzyme transition state configurations directly. It is becoming increasingly apparent that molecular simulation has a vital role to play in elucidating the complex processes involved in these outstanding natural catalysts. From the perspective of practical applications, modelling techniques that can shed new light on enzyme-catalysed reactions can then help to contribute toward the design of new drugs or the development of novel industrial catalysts via biomimetic approaches. The concept of using atomistic simulations to model enzyme-catalysed reactions, starting from the first pioneering works of Warshel [28] and Scheiner [29], has risen to prominence in recent years and is now at the point where the field of computational enzymology has securely laid foundations [30, 31]. However, there still remains little consensus about the ideal methodology to perform calculations on an enzyme of choice. Whilst it is outside the scope of this dissertation to discuss the matter in detail, a more comprehensive discussion of methods currently used, along with their associated advantages and disadvantages, can be found in an elegant recent review by Lonsdale [32]. Elucidation of the origins of enzyme catalysis involves understanding the origin of the difference between the uncatalysed activation barrier and the activation barrier in the protein, along with the associated enzyme mechanisms. The primary focus is on the factor that governs the reduction of the activation barrier of the chemical step. This is, of course, a question of energetics. One of the main objectives of this dissertation is to provide energies of activation and reaction for the rearrangement of chorismate to prephenate, both in the presence of the *Bacillus subtilis* chorismate mutase (CM) enzyme and also the uncatalysed equivalent reaction in water.

## 2.4 Summary

This chapter has outlined the biological concepts relevant to the investigations presented in this dissertation. Starting from the fundamental building blocks of nature, the amino acids, the discussion moved on to show how these component parts form larger polypeptide chains. The types of secondary structure motifs that these long chains fold into was then outlined. One of the essential dogmas of biology is that structure informs function. Once the polypeptide chains discussed fold into their correct structures they can then perform a variety of functions. One such function is to catalyse reactions that would otherwise take too long to be of biological relevance, demonstrating how enzymes are critical for life. The next chapter will discuss the computational methods used in this work to study proteins and the tests required to ensure that these methods accurately describe the physical properties we know are crucial for correctly describing the biochemistry of these systems. Chapter 4 will investigate the properties of small molecules that can adopt torsion angles, which correspond to those outlined in this chapter, to match those of amino acids that form extended  $\alpha$ -helices and  $\beta$ -sheets discussed here. If these properties can be reproduced using the ONETEP and OPTIM codes, discussed in the following chapter, then

the resulting simulations performed on larger systems in Chaps. 5 and 6 can be trusted. Transition state stabilisation, a key process in reducing the activation barrier, occurs, partially, as a result of efficient overlap of electron orbitals between the residues in the enzyme active site. The following chapter will outline accurate and efficient methods for optimising the ionic and electronic structure of enzyme systems and discuss how the bonding interactions between active-site residues and a substrate can be probed. These interactions in the active site of the CM enzyme can be analysed in detail, and this will be investigated in Chap. 6.

## References

1. F. Sanger, H. Tuppy, *Biochem. J.* **49**, 463 (1951)
2. B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Essential Cell Biology*, 3rd edn. (Garland Science, 2010)
3. *Structural information (determined by microwave spectroscopy) from CRC Handbook of Chemistry and Physics*, 88th edn (CRC Press, 2007)
4. D.J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003)
5. W.T. Astbury, *Trans. Faraday Soc.* **29**, 193 (1933)
6. L. Pauling, R.B. Corey, H.R. Branson, *Proc. Natl. Acad. Sci. USA* **37**, 205 (1951)
7. A. Bairoch, *Nucleic Acids Res.* **28**, 304 (2000)
8. E. Fischer, *Ber. Chem. Ges. Berl.* **27**, 2985 (1894)
9. H. Eyring, M. Polanyi, *Z. Phys. Chem. Abt. B* **12**, 279 (1931)
10. M.G. Evans, M. Polanyi, *Trans. Faraday Soc.* **31**, 875 (1935)
11. L. Pauling, *Am. Sci.* **36**, 50 (1948)
12. M.M. Mader, P.A. Bartlett, *Chem. Rev.* **97**, 1281 (1997)
13. R.A. Copeland, J.P. Davis, G.A. Cain, W.J. Pitts, R.L. Magolda, *Biochem.* **35**, 1270 (1996)
14. V.L. Schramm, *Annu. Rev. Biochem.* **80**, 703 (2011)
15. P. Kollman, B. Kuhn, M. Peräkylä, *J. Phys. Chem. B* **106**, 1537 (2002)
16. G. Hou, G. Hou, Q. Cui, *J. Am. Chem. Soc.* **134**, 229 (2011)
17. A. Fersht, *Enzyme Structure and Mechanism* (W. H. Freeman and Company, New York, 1985)
18. A. Radzicka, R. Wolfenden, *Science* **267**, 90 (1995)
19. N.J. Mulder, P. Kersey, M. Pruess, R. Apweiler, *Mol. Biotechnol.* **38**, 165 (2008)
20. J.P. Overington, B. Al-Lazikani, A.L. Hopkins, *Nat. Rev. Drug Discov.* **5**, 993 (2006)
21. A.L. Hopkins, C.R. Groom, *Nature Rev. Drug Discov.* **1**, 727 (2002)
22. A.J. Mulholland, *Drug Discov. Today* **10**, 1393 (2005)
23. M. Pirmohamed, B.K. Park, *Toxicology* **192**, 23 (2003)
24. L. Ridder, A.J. Mulholland, *Curr. Top. Med. Chem.* **3**, 1241 (2003)
25. M. Dantus, R.M. Bowman, J.S. Baskin, A.H. Zewail, *Chem. Phys. Lett.* **159**, 402 (1989)
26. S. Pedersen, L. Bañares, A.H. Zewail, *J. Chem. Phys.* **97**, 8801 (1992)
27. J.C. Polanyi, A.H. Zewail, *Acc. Chem. Res.* **28**, 1992 (1995)
28. A. Warshel, M. Levitt, *J. Mol. Biol.* **103**, 227 (1976)
29. S. Scheiner, W.N. Lipscomb, *Proc. Natl. Acad. Sci. USA* **73**, 432 (1976)
30. M.A. Cunningham, P.A. Bash, *Biochimie* **79**, 687 (1997)
31. T.C. Bruice, K. Kahn, *Curr. Opin. Chem. Biol.* **4**, 540 (2000)
32. R. Lonsdale, J.N. Harvey, J. Adrian, *Mulholland. Chem. Soc. Rev.* **41**, 3025 (2012)

Large-Scale Quantum-Mechanical Enzymology

Lever, G.

2015, XVII, 148 p. 30 illus., 12 illus. in color., Hardcover

ISBN: 978-3-319-19350-2