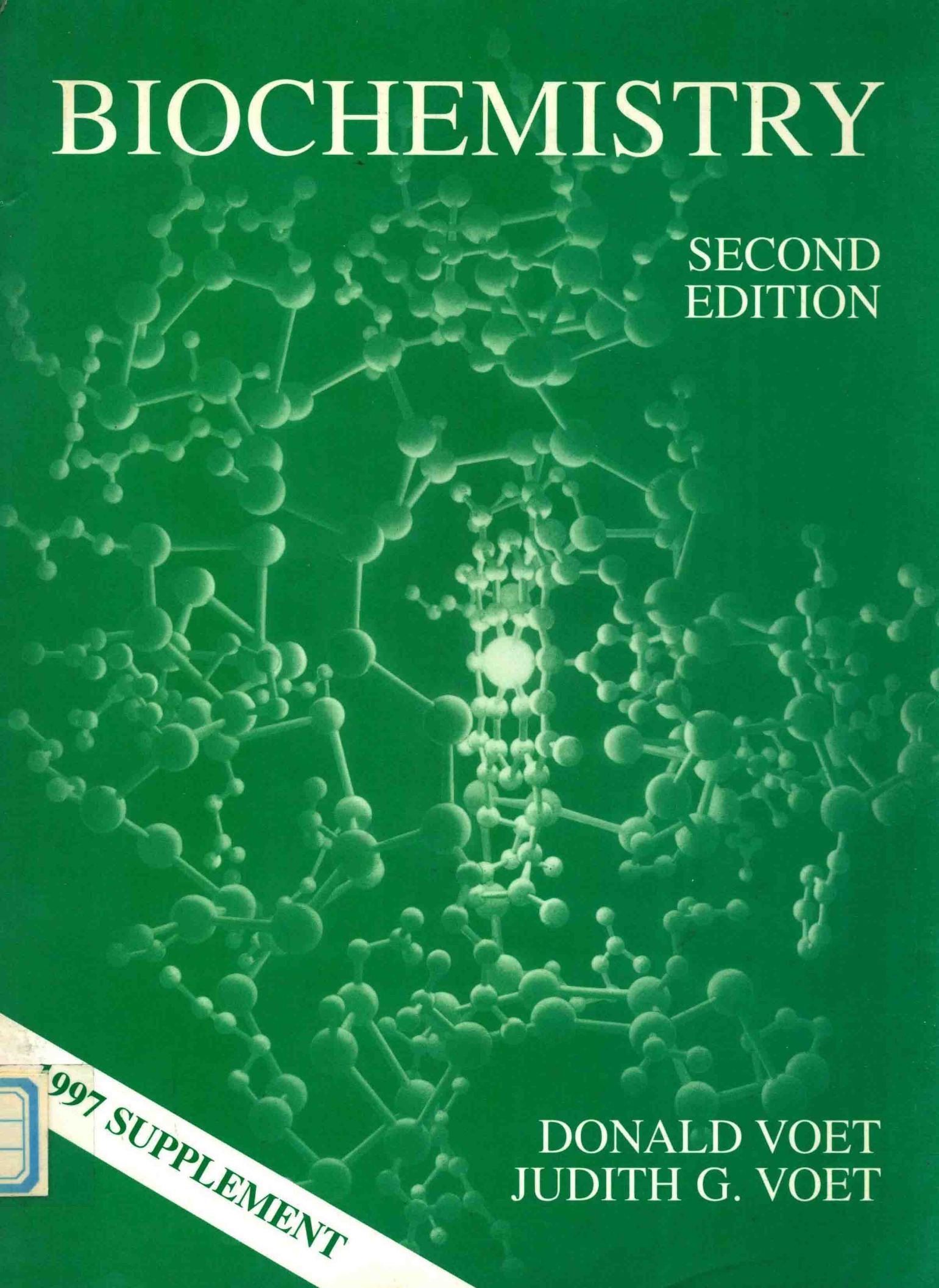
BIOCHEMISTRY

SECOND EDITION



1997 SUPPLEMENT

DONALD VOET
JUDITH G. VOET

DONALD VOET
University of Pennsylvania
JUDITH G. VOET
Swarthmore College

BIOCHEMISTRY

SECOND EDITION

1997 Supplement

JOHN WILEY & SONS, INC.
New York Chichester Weinheim
Brisbane Singapore Toronto

Cover art © *Irving Geis*

ISBN 0-471-19412-3

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Printed and bound by Victor Graphics, Inc.

INTRODUCTION

This is the second annual Supplement to *Biochemistry* (2nd ed.) by Donald Voet and Judith G. Voet. The biochemical literature covered in this Supplement extends from around April, 1996, to around April, 1997. During this ~12 month time period, the biochemical literature has continued its explosive growth. It is therefore increasingly important for the student and teacher alike to keep up with the literature. The annual Supplements to *Biochemistry* should be taken as a guide for doing so.

As with previous supplements to *Biochemistry,* this Supplement is keyed to the textbook in that we refer to new advances in terms of the textbook sections in which they would logically fit. Since space limitations permit only the most cursory discussions of these topics, the interested reader should consult the pertinent references provided at the end of each discussion. References that are not preceded by a discussion or which are placed under the subheading "Additional References" provide up-to-date coverage of subjects discussed in the corresponding section of the textbook. Note that many of these references are followed by a capsule synopsis.

The reader should note that two types of ancillary materials to accompany *Biochemistry* (2nd ed.) are available:

- A set of two CD-ROMs containing nearly all of the figures in the Second Edition, one in a form suitable for projection and the other in a form suitable for printing (ISBN 0-471-05882-3).

- A diskette containing a series of kinemages in 22 Exercises for the interactive display of various proteins and nucleic acids (ISBN 0-471-13558-5 for IBM-compatible computers and ISBN 0-471-13559-3 for Macintosh computers). We plan to develop several additional Exercises, which will be made available over the World Wide Web at http://www.wiley.com/college/voet

We are grateful to Irving Geis for providing the cover illustration of this Supplement, a monochrome version of his painting of horse heart cytochrome *c* illuminated by its iron atom, which is the front cover illustration of *Biochemistry* (2nd ed.).

If you purchased this Supplement or *Biochemistry* (2nd ed.) directly from John Wiley & Sons, Inc., your subscription for this update service has been recorded. If, however, you purchased this Supplement from a bookstore and wish to receive future Supplements billed separately, with a 30-day examination review, please send your name, company name (if applicable), address, and the notation "SUBSCRIPTION SERVICE FOR VOET/VOET: BIOCHEMISTRY, Second Edition" to: Supplement Department, John Wiley & Sons, Inc., One Wiley Drive, Somerset, NJ 08875, Tel. 1-800-225-5945, FAX: 908-302-2300. For customers outside the United States, please contact the Wiley office nearest you at the addresses given on the bottom of page vi of the textbook.

Donald Voet
Judith G. Voet

Chapter 1

LIFE

2. EUKARYOTES

Gupta, R.S. and Golding, G.B., The origin of the eukaryotic cell, *Trends Biochem. Sci.* **21,** 166–171 (1996).

4. THE ORIGIN OF LIFE

(a) Life May Have Been Present On Earth for at Least 3.85 Billion Years

The earliest known microfossils, which are ~3.5 billion years old, were generated by organisms resembling modern bacteria. Hence life must have arisen well before this time. However, the known examples of sedimentary rock that are more than ~3.5 billion years old have been subject to such extensive metamorphic forces (500°C and 5000 atm.) that any microfossils they contained would have been obliterated. Consequently, evidence for life on earth at the time these older rocks were formed can only be demonstrated through the geochemical analysis of minerals that are resistant to metamorphism.

In modern oceans, the biological generation of organic molecules and carbonates from CO_2 by photosynthesis and precipitation results, through physico-chemical effects, in the depletion of the amount of the ^{13}C present by 2.5% and thus in a decrease of the $^{13}C/^{12}C$ ratio relative to carbon of abiotic origin. Similarly decreased $^{13}C/^{12}C$ ratios persist in rocks up to 3.5 billion years old.

The oldest known sedimentary rocks, which occur in western Greenland, are at least 3.85 billion years old. The use of an ion-microprobe mass spectrometer (an instrument that can determine the isotopic compositions of granules as small as 20 μm) has established that the carbonaceous inclusions within grains of apatite (basic calcium phosphate) from these rocks have $^{13}C/^{12}C$ ratios similar to the above values. Since there is no known abiotic process that can both generate such isotopically light carbon and selectively incorporate it into apatite grains, it seems likely that these carbonaceous inclusions are of biological origin. If this is true (and it is disputed), then life on earth must have existed at the time these sedimentary rocks were laid down, that is, at least 3.85 billion years ago.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P., and Friend, C.R.L., Evidence for life on earth before 3,800 million years ago, *Nature* **384,** 55–59 (1996).

Holland, H.D., Evidence for life on earth more than 3850 million years ago, *Science* **275,** 38–39 (1997).

(b) Additional References

Lazcano, A. and Miller, S.L., The origin and early evolution of life: Prebiotic chemistry, the pre-RNA world, and time, *Cell* **85,** 793–798 (1996).

Chapter 5

TECHNIQUES OF PROTEIN PURIFICATION

Doonan, S. (Ed.), *Protein Purification Protocols, Methods in Molecular Biology,* v. 59, Humana Press (1996).

Bollag, D.M., Rozycki, D., and Edelstein, S.J., *Protein Methods* (2nd ed.), Wiley-Liss (1996).

Rosenberg, I.M., *Protein Analysis and Purification: Benchtop Techniques,* Birkhauser (1996).

Schuster, T.M. and Toedt, T.M., New revolutions in the evolution of anylytical ultracentrifugation, *Curr. Opin. Struct. Biol.* **6,** 650–658 (1996).

Chapter 6

COVALENT STRUCTURES OF PROTEINS

Chapman, J.R. (Ed.), *Protein and Peptide Analysis by Mass Spectrometry,* Humana Press (1996).

Fitxgerald, M.C. and Siuzdak, G., Biochemical mass spectrometry: worth the wait, *Chem. Biol.* **3,** 707–715 (1996).

Havel, H.A. (Ed.), *Spectroscopic Methods for Determining Protein Structure,* VCH Publishing (1996).

Hubbard, T.J.P., New horizons in sequence analysis, *Curr. Opin. Struct. Biol.* **7,** 190–193 (1997). [Discusses sophisticated methods for elucidating evolutionary relationships between polypeptides that have little apparent sequence similarity.]

Chapter 7

THREE-DIMENSIONAL STRUCTURES OF PROTEINS

1. SECONDARY STRUCTURE

(a) The Significance of Disallowed Ramachandran Conformations

In a set of 110 high resolution, non-homologous crystal structures containing 18,708 non-Gly residues, only 66 residues (0.4%) lie significantly outside the allowed regions of the Ramachandran diagram. The (ϕ, ψ) angles of these distorted residues cluster into six distinct regions of the Ramachandran diagram, thereby suggesting that certain backbone distortions are preferred in constrained structures. Indeed, 33 of these disallowed conformations are conserved in homologous proteins whose structures have been determined. The amino acid residues that most frequently adopt these disallowed conformations are the small, polar/charged, hydrogen bond-forming residues Asn, Asp, Ser, Thr, and His; few are bulky and hydrophobic. In all of these

distorted residues, the backbone bond angles about each C_α atom depart significantly from their ideal values in a way that eliminates or reduces the number of short van der Waals contacts that would otherwise occur.

Gunasekaran, K., Ramakrishnan, C., and Balaram, P., Disallowed Ramachandran conformations of amino acid residues in protein structures, *J. Mol. Biol.* **264,** 191–198 (1996).

(b) Additional References

Kohn, W.D., Mant, C.T., and Hodges, R.S., α-Helical protein assembly, *J. Biol. Chem.* **272,** 2583–2586 (1997).

Lupas, A., Coiled coils: new structures and functions, *Trends Biochem. Sci.* **21,** 375–382 (1996).

2. FIBROUS PROTEINS

Kadler, K.E., Holmes, D.F., Trotter, J.A., and Chapman, J.A., Collagen fibril formation, *Biochem. J.*, **316,** 1–11 (1996).

3. GLOBULAR PROTEINS

Mozzarelli, A. and Rossi, G.L., Protein function in the crystal, *Annu. Rev. Biophys. Biomol. Struct.* **25,** 343–365 (1996).

4. PROTEIN STABILITY

(a) Structural Features Stabilizing Halophilic Proteins

Most proteins precipitate and hence cease to function at salt concentrations >1*M*. This phenomenon, which is known as salting out (Section 5-2A), is primarily due to the ability of inorganic ions to bind and thereby sequester the water molecules that proteins require for solvation. Yet archaebacteria that live in such hypersaline environments as the Great Salt Lake and the Dead Sea accumulate inorganic ions such that their intracellular salt concentrations exceed those of their environments (and hence are supersaturated). In fact, **halophilic** (salt-loving) enzymes, which catalyze the same reactions as do their nonhalophilic homologs, require salt concentrations in the range 1 to 4*M* for stability and activity. How do halophilic proteins differ from their "normal" counterparts?

The X-ray structures of two halophilic enzymes are presently known: those of a malate dehydrogenase (MDH, a citric acid cycle enzyme; Section 19-3H) and a [2Fe–2S] ferredoxin (a redox enzyme; Section 20-2C), both from *Haloarcula marismortui*, an archaebacterium that thrives in the Dead Sea (the saltiest body of water on Earth, which contains 4*M* KCl and many other salts). Both of these proteins have a far greater number of acidic residues on their surfaces than do their nonhalophilic homologs. Thus, the 128-residue monomeric *H. marismortui* ferredoxin (HmFd) has 34 Asp and Glu residues but only 6 basic (Arg and Lys) residues for a net charge of –28 (which gives it the highest negative surface charge density of any known protein, -4.3×10^{-3} Å^{-2}), whereas the 33% identical [2Fe–2S] Fd from the cyanobacterium *Anabaena* has a net charge of –13. Similarly, the 303-residue subunits of the homotetrameric HmMDH each contain 62 Asp and Glu residues for a net charge of –39, whereas a subunit of the ~37% identical lactate dehydrogenase (LDH; Section 16-3A) from dogfish has a net charge of +16.

Why do halophilic proteins have such an excess of acidic residues? It has been demonstrated that carboxylate groups have a far greater capacity to bind water molecules than do other protein groups. Thus it appears that these acidic residues permit halophilic proteins to effectively compete with inorganic ions for water of solvation and thereby remain in solution.

Dym, O., Mevarech, M., and Sussman, J.L., Structural features that stabilize halophilic malate dehydrogenase from an archeabacterium, *Science* **267,** 1344–1346 (1995).

Frolow, F., Harel, M., Sussman, J.L., Mevarech, M, and Shoham, M., Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin, *Nature Struct. Biol.* **3,** 452–458 (1996).

(b) The X-ray Structures, Instability, and Aggregation of Human Lysozyme Variants Causing Amyloid Fibrillogenesis

A number of ultimately fatal diseases, including **systemic amyloidosis,** Alzheimer's disease, and transmissible spongiform encephalopathies (e.g., scrapie; Section 32-5B), are associated with hereditary variants of normally monomeric proteins that are deposited in tissues as **amyloid fibrils.** Remarkably, the core structures of all amyloid fibrils, no matter what the protein from which they are formed, consists of β sheets whose polypeptide strands are perpendicular to the fiber axis. There are two naturally occurring variants of human lysozyme known to cause autosomal dominant hereditary amyloidosis, Lys56Thr and Asp67His. The resulting amyloid fibrils consist exclusively of the variant proteins.

The X-ray structures of the two amyloidogenic human lysozyme variants resemble that of the wild-type enzyme (which is 60% identical in sequence to hen egg white lysozyme; Section 14-2). Both have the four correct, intact disulfide bonds. However, the replacement of Asp 57 by His interrupts a network of hydrogen bonds that stabilizes the domain containing the β sheet, thereby causing a distortion of the active site and a movement of the backbone atoms near residues 48 and 70 by up to 11 Å. The replacement of Ile 56 by Thr insinuates a hydrophilic residue in a critical hydrophobic interface that links the protein's α and β domains, although it causes only subtle changes in the protein structure.

The melting temperatures (T_m) of both variants are at least 10°C less than those of wild-type human lysozyme, and both variants eventually lose all enzymatic activity when incubated at physiological temperature and pH (37°C and 7.4), conditions under which wild-type lysozyme remains fully active. The amyloidogenic lysozyme variants also aggregate on heating *in vitro* and a variety of physical measurements indicate that, in doing so, they form amyloid-like fibrils. In hydrogen exchange experiments (Section 8-2), wild-type lysozyme strongly protects 55 protons from exchange with D_2O under conditions (37°C and pH 5) that these protons are essentially unprotected in the amyloidogenic variants, thereby confirming that the two mutations greatly loosen the native protein's tertiary structure. This suggests that the partially-folded, aggregation-prone forms are in dynamic equilibrium with the native conformation, even under the conditions that the native state is thermodynamically stable. It has therefore been proposed that lysozyme fibrillogenesis is initiated by the association of the β domains of two partially unfolded lysozyme variants to form a more extensive β sheet. This would provide a template or nucleus for the recruitment of additional polypeptide chains to form the growing fibril in a process that may involve the conformational conversion of α helices to β strands. This autocatalytic refolding process may be a general mechanism for amyloid fibrillogenesis.

Booth, D.R., *et al.*, Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis, *Nature* **385,** 787–793 (1997).

Funahashi, J., Takano, K., Ogasahara, K., Yamagata, Y., and Yutani, K., The structure, stability and folding process of amyloidogenic mutant human lysozyme, *J. Biochem.* **120,** 1216–1223 (1996).

(c) Additional References

Derewenda, Z.S., Lee, L., and Derewenda, U., The occurrence of C—H···O hydrogen bonds in proteins, *J. Mol. Biol.* **252,** 248–262 (1995).

Ladbury, J.E., Just add water! The effect of water on the specificity of protein–ligand binding sites and its potential application to drug design, *Chem. Biol.* **3,** 973–980 (1996).

5. QUARTERNARY STRUCTURE

Jones, S. and Thornton, J.M., Principles of protein–protein interactions, *Proc. Natl. Acad. Sci.* **93,** 13–20 (1996). [Describes the characteristics of subunit–subunit interfaces based on a survey of known structures of homodimeric proteins, heterodimeric proteins, enzyme–protein inhibitor complexes, and antibody–protein complexes.]

Chapter 8

PROTEIN FOLDING, DYNAMICS, AND STRUCTURAL EVOLUTION

1. PROTEIN FOLDING: THEORY AND EXPERIMENT

(a) Secondary Structure Is Context-Dependent

The structure of a native protein is determined by its amino acid sequence. But, to what extent is the conformation of a given polypeptide segment influenced by the surrounding protein?

Protein GB1, the 56-residue IgG-binding domain of Streptoccocal **protein G**, binds to the Fc region of immunoglobulin G (IgG), thereby helping the bacterium evade the host's immunological defenses (IgG is discussed in Section 34-2B). GB1's NMR structure reveals that this protein, which lacks disulfide bonds, consists of a long α helix lying across a 4-stranded mixed β sheet. Through site-directed mutagenesis, the 11-residue "chameleon" sequence AWTVEKAFKTF was used to replace either GB1's α helix residues 23 to 33 (AATAEKFVFQY in GB1; a 7-residue change) to form Chm-α or its β sheet residues 42 to 52 (EWTYDDATKTF in GB1, which forms segments of two β strands connected by a β-turn; a 5-residue change) to form Chm-β. Both Chm-α and Chm-β display reversible thermal unfolding typical of compact single-domain globular proteins and their NMR spectra indicate that each assumes a structure similar to that of native GB1. Yet, NMR measurements also demonstrate that the isolated chameleon polypeptide (Ac-AWTVEKAFKTF-NH_2, where Ac is acetyl) is unfolded in solution, which suggests that this sequence has no strong preference for either an α helix or a β sheet conformation. Evidently, the information specifying α helix or β sheet secondary structures can be entirely nonlocal, that is, context-dependent effects are important in protein folding.

Minor, D.L., Jr. and Kim, P.S., Context-dependent secondary structure formation of a designed protein sequence, *Nature* **380,** 730–734 (1996).

(b) Preferred Structures in Proteins

Under physiological conditions, a polypeptide's 1-dimensional sequence specifies its 3-dimensional structure. Yet, most mutations, even those consisting of extensive insertions or deletions, only locally perturb a protein's structure. Indeed, proteins with no apparent sequence similarity may have quite similar folds (e.g., the α/β-barrel proteins; Section 16-2E).

Insights into why polypeptides can have unique structures and why these structures are relatively insensitive to sequence are provided by a surprisingly crude model. In this model, a polypeptide is represented by beads on a string that are constrained to occupy all points of a regular

lattice such as a cubic lattice of $3 \times 3 \times 3$ points (so that here the polypeptide consists of 27 residues). Only two types of residues are allowed, those that are "hydrophobic" (H) and those that are "polar" (P). Thus a sequence is defined by the type of residue, H or P, at each of its positions. Contacting residues therefore can have one of three types of interaction energies, E_{HH}, E_{PP} and E_{HP}, where $E_{PP} > E_{HP} > E_{HH}$, so that configurations in which H residues are buried (out of contact with water) have lower energies. The overall energy of a given sequence with a particular structure is defined as the sum of the interactions energies of all its contacting residues that are not adjacent in sequence.

There are 51,704 distinguishable ways that a 27-residue polypeptide can be folded into a $3 \times 3 \times 3$ lattice. The computation of the overall energies of all of these structures for each of the 2^{27} different sequences (for a total of 6.9×10^{12} combinations of structure and sequence) yielded intriguing results. A structure that is the lowest energy form of a given sequence is said to be "designed" by that sequence. Certain structures are designed by a large number of sequences, whereas other structures are designed by only a few or no sequences. Highly designable structures have characteristic "secondary structures" consisting of straight runs of residues folded in a regular manner and have lower overall energies than do other structures. Moreover, in the 60 most highly designable structures (0.12% of the 51,704 possible structures), the energy gap between the lowest energy structure of a given sequence and that of its next highest energy structure is particularly large. Thus, the most highly designable structures are also the most stable.

The comparison of the sequences forming the most highly designable structures is also enlightening. Some of the residue sites are highly mutable, whereas others are highly conserved. The conserved residues are those with either the smallest or the largest number of sites exposed to the exterior. Nevertheless, sequences with no apparent sequence similarity can design the same structure.

The properties of this crude lattice model are remarkably similar to those of real proteins. Thus, it might be expected that highly designable protein structures (those to which many polypeptides fold) are more stable than are less designable structures. If this is true, it seems likely that random selection in the primordial stages of protein evolution chose the more highly designable structures because these structures are likely to be the least sensitive to mutation.

Li, H., Helling, R., Tang, C., and Wingreen, N., Emergence of preferred structures in a simple model of protein folding, *Science* **273,** 666–669 (1996).

(c) A "New View" of Protein Folding Based on Submillisecond Observations

The classical view of protein folding, which is based on measurements spanning times of seconds or more, is that proteins fold through a series of well defined intermediates that may or may not be on the direct pathway between the fully unfolded polypeptide and the native protein. However, the advent of experimental methods for observing how proteins fold on the submillisecond time scale supports a rather different view of how proteins fold, called **landscape theory,** in which the folding process is envisioned to occur on an energy surface or landscape.

The techniques that appear to be particularly effective at initiating protein folding in a way that it can be monitored in the submillisecond regime include ultrarapid mixing so as to dilute a denaturant or change the pH and, in a technique known as temperature jump, the rapid heating, via an infrared laser pulse, of a cold denatured protein (since $\Delta G = \Delta H - T\Delta S$, proteins are unstable, that is denature, when $T < \Delta H/\Delta S$; for many proteins, solution conditions can be found for which this temperature is >0°C). Protein folding is then monitored, for example, by measuring the changes in the fluorescence or circular dichroism (CD) spectrum of the folding protein.

The energy states available to a polypeptide may be represented by an energy surface. The horizontal coordinates of any point on this surface are indicative of a particular conformation of the polypeptide and the vertical coordinate represents its internal free energy in that conformation. The above-described measurements indicate that the energy surface of a folding polypeptide is funnel-shaped with the native state represented by the bottom of the funnel, the global free energy minimum. The width of the funnel at any particular height (free energy) above the native state is

indicative of the number of conformational states with that free energy, that is, the entropy of the polypeptide.

Polypeptides fold via a series of conformational adjustments that reduce their free energy and entropy until the native state is reached. Since a collection of unfolded polypeptides all have different conformations (have different positions on the folding funnel), they cannot follow precisely the same pathway in folding to the native state. If the polypeptide actually folded to its native state via a random conformational search as Levinthal conjectured (*p.* 194), the folding funnel would resemble a flat disk with a single small hole, much like the surface of a golf course. Thus, it would take an enormously long time for a polypeptide to achieve the native state (fall in the hole) via a random conformational search (by rolling about at random on the surface of the golf course).

The folding funnel of a protein that follows the classical view of protein folding would have a deep radial groove in its disk-like surface that slopes towards the hole representing the native state. The extent of the conformational search to randomly find this groove would be much reduced relative to the Levinthal model so that such a polypeptide would readily fold to its native state. However, the conformational search for the pathway (groove) leading to the native state would still take time so that the polypeptide would require perhaps several seconds to start down the folding pathway.

The above-described measurements reveal that many polypeptides form significant native-like structure within fractions of a millisecond after commencing folding. This indicates that the folding funnel is, in fact, funnel-shaped, that is, it tends to slope towards the native conformation at all points. Thus, the various pathways followed by initially unfolded polypeptides in folding to their native state are analogous to the various trajectories that could be taken by skiers initially distributed around the top of a bowl-shaped valley to reach the valley's lowest point. Apparently, there is no single pathway or closely related set of pathways that a polypeptide must follow in folding to its native conformation.

The foregoing does not imply that the surface of the folding funnel is necessarily smooth. Indeed, landscape theory suggests that this energy surface has a relatively rugged landscape, that is, it has many local minima. Consequently, in following any particular folding pathway, a polypeptide is likely to become trapped in a local minimum until it randomly acquires sufficient thermal energy to overcome this kinetic barrier and continue the folding process. Thus, in landscape theory, the transition states that govern the rate of folding are not specific structures as they are assumed to be in the classical theory of protein folding but, rather, are ensembles of structures.

Dill, K.A., and Chan, H.S., From Levinthal pathways to funnels, *Nature Struct. Biol.* **4,** 10–19 (1997). [A highly readable account of landscape theory and its implications.]

Eaton, W.A., Moñoz, V., Thompson, P.A, Chan, C.-K., and Hofrichter, J., Submillisecond kinetics of protein folding, *Curr. Opin. Struct. Biol.* **7,** 10–14 (1997).

Eaton, W.A., Thompson, P.A, Chan, C.-K., Hagen, S.J., and Hofrichter, J., Fast events in protein folding, *Structure* **4,** 1133–1139 (1996).

Wolynes, P.G., Luthey-Schulten, Z., and Onuchic, J.N., Fast-folding experiments and the topography of protein folding energy landscapes, *Chem. Biol.* **3,** 425–432 (1996).

(d) X-Ray Structures of the DnaK Peptide-Binding Unit in Complex with a Peptide and of the DnaK ATPase in Complex with GrpE

The Hsp70 family of molecular chaperone proteins facilitate the proper folding of unfolded proteins (e.g., polypeptides emerging from a ribosome), as well as the unfolding of proteins preparatory to their transport through membranes (Section 8-1C). **DnaK,** the *E. coli* Hsp70 protein (so-named because it was initially implicated in the replication of bacteriophage λ DNA in *E. coli*) is 40 to 50% identical in sequence to its eukaryotic homologs. DnaK, as well as other

Hsp70 proteins, is a monomeric protein that consists of an N-terminal 42-kD ATPase that has no peptide-binding affinity and a 25-kD C-terminal unit that binds predominantly hydrophobic peptides. DnaK acts in concert with two other heat shock proteins: **DnaJ,** a member of the **Hsp40** family of molecular chaperones, which binds to denatured proteins and certain native proteins and probably to DnaK's peptide-binding unit; and **GrpE,** a nucleotide exchange factor that functions analogously to the guanine nucleotide releasing factors (GRFs) associated with many GTP-binding proteins (Sections 30-3D and 34-4B). The binding and release of peptides by DnaK is governed by conformational changes induced by the binding and hydrolysis of ATP. This appears to occur as follows: ATP-bound DnaK rapidly binds and releases its substrate peptides. DnaJ stimulates the hydrolysis of this ATP, yielding ADP-bound DnaK, which binds peptide tightly. Peptide release requires the GrpE-catalyzed dissociation of ADP, a process that is rapidly followed by ATP rebinding.

The X-ray structure has been determined of DnaK's C-terminal peptide-binding unit (residues 389-607 of the 638-residue protein) in complex with a tight-binding heptapeptide of sequence NRLLLTG. The protein consists of two domains: an N-terminal β sandwich consisting of two 4-stranded antiparallel β sheets; and a C-terminal domain consisting mainly of five α helices, three of which form an antiparallel bundle that lies on top of the β sandwich and is approximately parallel to its β strands. The substrate heptapeptide is bound in an extended conformation between the 3-helix bundle and the β sandwich in a channel formed by two of the loops connecting the β strands. It is oriented approximately normal to the β strands. The heptapeptide's central Leu sidechains interact with the protein mainly through hydrophobic interactions, thereby accounting for DnaK's preference for binding predominantly hydrophobic peptides. The substrate-binding site is largely occluded by the protein so that it would be quite difficult for even a short peptide to thread its way into or out of this binding site. Thus it seems likely that peptide binding and release requires a conformational change in which the α helical lid covering the substrate-binding channel lifts away to expose this channel. Indeed, the X-ray structure of a second crystal form of this same complex reveals just such a conformational change relative to the complex in the first crystal form, although of insufficient magnitude to fully uncap the substrate-binding channel. It is therefore postulated that this conformational change is incipient to the full-scale change required for peptide exchange. Thus it seems likely that the GrpE-facilitated exchange of DnaK's bound ADP for ATP induces such a conformational change in its substrate-binding unit.

The X-ray structure of DnaK's ATPase fragment (residues 1-388) in complex with a homodimer of GrpE (residues 34-197 of its 197-residue subunits) has also been determined. The ATPase forms a deeply clefted structure that is largely superimposable on the X-ray structures of the 55% identical ATPase fragment of the bovine brain chaperone protein **Hsc70** (Hsc for *h*eat *s*hock *c*ognate) in complex with ADP and of the ATPase domain of human Hsp70 in complex with ADP and P_i. ADP binds to the Hsc70 and Hsp70 ATPases at the bottom of this cleft. The tertiary structures of the nucleotide-binding cores of the Hsc70, Hsp70, and DnaK ATPases are similar to that of the glycolytic enzyme hexokinase (Section 16-2A), suggesting that the phosphotransferase functions and the substrate-induced conformational changes of these proteins occur via similar mechanisms.

The N-terminal segment of each GrpE monomer consists of a 19-turn (~100-Å-long) α helix, which associates in parallel with the N-terminal helix of the dimer's other GrpE subunit to form a long two-helix bundle. Each chain then turns back on itself in a second α helix, so that the 2-helix bundle terminates in a 4-helix bundle.The chains each end in a small β sheet domain that extends outwards from the base of the 4-helix bundle. Consequently, the GrpE dimer has an appearance resembling the Greek letter ψ with a greatly elongated stem. The DnaK ATPase domain binds asymmetrically along one side of the top portion of the ψ such that it contacts only one of GrpE's monomer units, thereby slightly distorting this monomer relative to the other. Many of these contacts are made across the top of the ATPase's binding cleft. Indeed, one lobe of this binding cleft is rotated by 14° relative to its position in the Hsc70 structure so as to open up the binding cleft and disrupt the ATP binding site. Thus GrpE-induced nucleotide exchange appears to simply involve the opening of DnaK's nucleotide-binding cleft. It is postulated that the N-terminal

tail of GrpE's 2-helix bundle (the stem of the elongated ψ), which projects well beyond its bound ATPase, interacts with DnaK's peptide-binding unit.

Flaherty, K.M., DeLuca-Flaherty, C., and McKay, D.B., Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein, *Nature* **346,** 623–628 (1990).

Harrison, C.J., Hayer-Hartl, M., Di Liberto, M., Hartl, F.-U., and Kuriyan, J., Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK, *Science* **276,** 431–435 (1997).

Ruddon, R.W. and Bedow, E., Assisted protein folding, *J. Biol. Chem.* **272,** 3125–3128 (1997).

Sriram, M., Osipiuk, J., Freeman, B.C., Morimoto, R.I., and Joachimiak, A., Human Hsp70 molecular chaparone binds two calcium ions within the ATPase domain, *Structure* **5,** 403–414 (1997). [The X-ray structure of the ATPase domain of human Hsp70.]

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E., and Hendrickson, W.A., Structural analysis of substrate binding by the molecular chaperone DnaK, *Science* **272,** 1606–1614 (1996).

(e) Concerted Movements in GroEL and its Complexes with GroES

Cycles of ATP turnover in GroEL–GroES complexes stimulate cycles of binding and release of the unfolded substrate protein, ultimately resulting in its folding to the native state. Through cryoelectron microscopy and three-dimensional image reconstruction, the structures of GroEL and its complexes with GroES in the presence of ATP, ADP, and the nonhydrolyzable ATP analog AMP-PNP have been visualized at 30-Å resolution.

Nucleotide binding to GroEL's equatorial domain induces large rotations of the GroEL apical domains, which contain the protein's GroES- and substrate protein-binding sites. In doing so, the complexes maintain their 7-fold rotational symmetry so that these structural changes are concerted and hence allosteric. These movements appear to twist the substrate binding sites, which face the protein's central channel, towards the subunit contact regions, thereby occluding these binding sites. The mechanical basis for the concerted nature of these movements is that untwisted domains would block the twisting of adjacent domains. In the GroEL–GroES complex, the GroEL heptameric ring to which GroES binds assumes a markedly different conformation from the GroEL heptameric ring not in contact with GroES. These domain movements are best appreciated by viewing successive conformational forms as movies, available on the World Wide Web at http://www.cryst.bbk.ac.uk/~ubcg16z/elmovies.html.

Roseman, A.M., Chen, S., White, H., Braig, K., and Saibil, H., The chaperonin ATPase cycle: Mechanism of allosteric switching and movements of substrate-binding domains in GroEL, *Cell* **87,** 241–251 (1996).

(f) Additional References

Aurora, R., Creamer, T.P., Srinivasan, S., and Rose G.D., Local interactions in protein folding: Lessons from the α-helix, *J. Biol. Chem.* **272,** 1413–1426 (1997).

Bohr, H. and Brunak, S. (Eds.), *Protein Folds,* CRC Press (1996).

Chan, H.S. and Dill, K.A., A simple model of chaperonin-mediated protein folding, *Proteins* **24,** 245–351 (1996). [Uses a 2-dimensional lattice model of protein folding to show how a chaperonin with a sticky surface can pull apart an incorrectly folded protein so that it can try to fold again. In this model, as has been experimentally observed for real proteins, the folding rate and the amount

of native protein produced can be greatly enhanced, or sometimes reduced, by varying the amino acid sequence, the chaperonin size, and the binding and ejection rates from the chaperonin.]

Ellis, R.J. (Ed.), *The Chaperonins,* Academic Press (1996).

Friesner, R.A. and Gunn, J.R., Computational studies of protein folding, *Annu. Rev. Biophys. Biomol. Struct.* **25,** 315–342 (1996).

Galat, A. and Metcalfe, S.M., Peptidylproline *cis/trans* isomerases, *Prog. Biophys. Molec. Biol.* **63,** 67–118 (1995).

Martin, J. and Hartl, F.U., Chaperone-assisted protein folding, *Curr. Opin. Struct. Biol.* **7,** 41–52 (1997); *and* Hartl, F.U., Molecular chaperones in cellular protein folding, *Nature* **381,** 571–580 (1996). [Authoritative reviews. They provide an updated model for the GroEL–GroES reaction cycle, which differs from that in Fig. 8-11 mainly in that the folding polypeptide does not pass between the two heptameric rings of GroEL subunits (the passage between these rings is occluded) and that these two rings bind and hydrolyze ATP at different stages of the reaction cycle.]

3. STRUCTURAL EVOLUTION

(a) Experimental Evidence that Evolution Favors Thermodynamically Stable Structures

The thermal denaturation temperature (T_m) of the native IgG-binding protein, GB1 (discussed in Section 8-1a, above), at pH 5.4 is 87.5°C, whereas the T_m's of two genetically engineered mutants, Chm-α and Chm-β, are 61.4°C and 39.2 °C. This indicates, in agreement with numerous similar observations in a variety of proteins, that the sequence of a polypeptide segment influences the stability of the particular fold in which it is embedded. The evolutionary significance of this phenomenon was investigated in GB1 by replacing the first three, four, or five residues of the previously discussed β-turn segment (DDATK, comprising GB1 residues 46-50) with all possible sequences of the same length. These residue replacements were made in wild-type GB1 and in two of its less stable but functional (at 25°C) variants, D22G and A26G. The β-turn segment that was varied is solvent-exposed and is distant from the residues involved in IgG binding.

The ability of up to 50,000 of these mutant GB1 proteins, selected at random, to bind IgG was determined. In the initially wild-type GB1 at 25°C, ~75% of the mutant proteins bind IgG. However, this fraction decreases to the range ~1 to 5% in the initially A26G and D22G variants and drops to below 1% at the T_m's of these variants (45°C and 65°C, respectively). In the variant proteins that bind IgG, the sequences of many of the inserts resemble either those of the wild-type β-turn or sequences with known preferences to form β-turns. Evidently, this β-turn is under evolutionary pressure to form a thermodynamically stable structure.

Zhou, H.X., Hoess, R.H., and DeGrado, W.F., *In vitro* evolution of thermodynamically stable turns, *Nature Struct. Biol.* **3,** 446–451 (1996).

(b) The Protein Universe Contains a Limited Number of Folds

An exhaustive all-on-all comparison of the ~5000 known protein structures reveals that "fold space" has five dominant regions. These so-called attractors contain folds of similar topologies that collectively comprise 40% of the known domains. In folded structures, these domains are buried, that is, not exposed to solvent, and hence are likely to form early in the folding process. On the basis of these observations and the rapidity with which new protein structures are being reported, it is predicted that we shall eventually know nearly all the folds that proteins are capable of forming.

Holm, L. and Sander, C., New structure — novel fold?, *Structure* **5,** 165–171 (1997); *and* Mapping the protein universe, *Science* **273,** 595–602 (1996).

(c) Additional References

Chung, S.Y. and Subbiah, S., A structural explanation for the twilight zone of protein sequence homology, *Structure* **4,** 1123–1127 (1996).

Murzin, A.G., Structural classification of proteins: new superfamilies, *Curr. Opin. Struct. Biol.* **6,** 386–394 (1996).

Chapter 9

HEMOGLOBIN: PROTEIN FUNCTION IN MICROCOSM

2. STRUCTURE AND MECHANISM

(a) Photolysis of the Carbon Monoxide Complex of Myoglobin Observed by Time-Resolved X-Ray Crystallography on the Nanosecond Timescale

Using modern laboratory equipment, it usually takes from hours to days to measure a set of X-ray diffraction intensities from protein crystals. Hence, the resulting structural model is indicative of the protein's average structure over this time period. However, the recent development of synchrotrons capable of generating enormously intense X-radiation together with techniques for making rapid intensity measurements now permits the measurement of an X-ray intensity data set on the nanosecond timescale. This capability has been demonstrated through the crystallographic study of the laser pulse-induced photodissociation of carbon monoxide from its complex with myoglobin (MbCO) and its subsequent rebinding.

Synchrotrons accelerate electrons around a circular track to near light speed. As a consequence, the electrons emit radiation within the plane of the circle. Although synchrotrons were first constructed for use as "atom smashers", it was soon realized that their so-called parasitic radiation, which is many orders of magnitude more intense than can be generated by ordinary laboratory equipment, could also be an invaluable tool in numerous applications. As a consequence, synchrotrons have been expressly designed to make use of their parasitic radiation. The electrons travel around the synchrotron's circular track in bunches rather than as a continuous distribution and hence the emitted radiation is pulsed rather than continuous.

CO, which is normally tightly bound by Mb, is photolytically displaced by visible light, a process which occupies less than a picosecond. The resulting structural changes in the Mb and the subsequent rebinding of the CO to the same Mb molecule (a process called geminate recombination) have been extensively studied by a variety of time-resolved spectroscopic techniques. However, in the absence of direct structural information, the interpretation of the spectral data is somewhat speculative.

The photolyis of a crystal of MbCO was initiated by a 7.5-ns laser pulse. Subsequent structural changes were then monitored using either a single 150-ps pulse of synchrotron-generated X-rays or a 940-ns train of these pulses. X-ray intensity data sets were obtained at six time delays between the laser and X-ray pulses: 4 ns, 1 μs, 7.5 μs, 50.5 μs, 350 μs and 1.9 ms. At 4 ns after the laser pulse, the CO is absent from its binding site on the heme but gradually rebinds to it until, by 350 μs, it is fully rebound. The X-ray data also show that at 4 ns, the Fe has moved from its position in MbCO, in the heme plane, towards the position it occupies in unliganded Mb, 0.32 Å

out of the heme plane towards its His ligand (the proximal His). This change decreases in parallel with the rebinding of the CO to the Mb. There are numerous small structural changes involving the residues surrounding the heme that also appear and disappear with time. Many of these are indicative of rearrangements of the residues forming the E and F helices. There is also a strong feature, which may represent the transient "docking site" for the photodissociated CO, that had been spectroscopically observed.

Thus, it has been clearly established that structural changes can be followed on the nanosecond timescale. In order to use this technique to monitor, for example, the structural changes that occur during an enzymatic reaction, all the enzyme molecules in a crystal must react in synchrony. Hence, methods must be developed to trigger enzymatic reactions, much like a laser pulse triggers the dissociation of all of the CO molecules in a crystal of MbCO.

Srajer, V., Teng, T., Ursby, T., Pradervand, C., Ren, Z., Adachi, S., Schildkamp, W., Bourgeois, D., Wulff, M., and Moffat, K., Photolysis of the carbon monoxide complex of myoglobin: Nanosecond time-resolved crystallography, *Science* **274,** 1726–1729 (1996).

(b) X-Ray Structures of T-State Oxyhemoglobin and R-State Deoxyhemoglobin

Under normal physiological conditions, deoxyhemoglobin is almost entirely in its T-state conformation, whereas fully oxygenated hemoglobin [$Hb(O_2)_4$] is almost entirely in its R state conformation. The switch from the T to the R state, which predominantly occurs at an intermediate state of oxygenation, is brought about by the buildup of strain due to tertiary structural changes in the subunits that become oxygenated and, conversely, the switch from the R to the T state is motivated by the tertiary structural strain in its deoxygenated subunits. Thus, a detailed understanding of the basis of hemoglobin's O_2-binding cooperativity requires knowledge of these tertiary structural changes. Heretofore, this information has been difficult to obtain because crystals of T-state Hb crack and become disordered on exposure to O_2 due to the weakness of the lattice forces holding the Hb molecules in the crystal, and similarly for deoxygenating crystals of R-state oxyHb.

Crystals of deoxyHb grown in the presence of inositol hexaphosphate (IHP, which binds to T-state Hb with high affinity) are stable when oxygenated at 4°C. The X-ray structure of these crystals show that O_2 binds to all four Hb subunits. In each α subunit, the Fe has moved from its position in T-state deoxyHb of 0.3 to 0.4 Å out of this plane on the side of the proximal His, into the heme plane, its position in R-state oxyHb. The proximal His is thereby displaced towards the heme, closer to its position in R-state Hb. The Fe—N_ε distance in T-state deoxyHb of 2.21 Å remains essentially unchanged so that the Fe—N_ε distance in T-state oxyHb is stretched by 0.3 to 0.4 Å relative to its 1.94-Å distance in R-state oxyHb. Nevertheless, the shift of the proximal His also slightly displaces its attached F helix towards the heme. The changes in the β subunits are similar to but generally smaller than those in the α subunits. There are substantial structural movements at the α_1–β_2 and α_2–β_1 interfaces (at which the major structural changes between Hb's T and R states occur) towards the R state quaternary structure, thereby weakening the T-state interface. All of these shifts are similar to but larger than those previously observed in T-state metHb and other partially liganded T-state forms, thereby providing clear evidence of heme–heme interactions in the T state.

Crystals of R-state deoxyHb were prepared by reducing crystals of metHb that had previously been chemically crosslinked. X-ray analysis of these crystals confirm that the Hb molecules are free of ligand but remain in the R state conformation. The structural shifts observed in the heme pockets and the α_1–β_2 and α_2–β_1 interfaces of R-state deoxyHb relative to those of R-state oxyHb are comparable in magnitude to those observed upon the deligation of T-state oxyHb, although they differ in direction. Nevertheless, whereas ligand binding in the T state appears to result in unfavorable stereochemistry around the heme–ligand complex, the R state appears to have a more plastic structure that can equally well accommodate liganded and ligand-free hemes. Thus, the greater stability of the deoxy T state over the deoxy R state is probably due to the formation of the salt bridge networks involving its C-terminal residues and the greater complementarity of its intersubunit contacts.

Paoli, M., Liddington, R., Tame, J., Wilkinson, A., and Dodson, G., Crystal structure of T state haemoglobin with oxygen bound at all four haems, *J. Mol. Biol.* **256,** 775–792 (1996).

Wilson, J., Phillips, K., and Luisi, B., The crystal structure of horse deoxyhaemoglobin trapped in the high-affinity (R) state, *J. Mol. Biol.* **264,** 734–756 (1996).

(c) A Test of the Role of the Proximal Histidines in Hemoglobin Cooperativity

In the Perutz mechanism for hemoglobin cooperativity, the movement of the Fe into the heme plane on liganding O_2 pulls the covalently attached His F8, the proximal His, after it. This, in turn, repositions the attached F helix, which results in a cascade of tertiary conformational shifts that ultimately trigger the T to R transition. In a test of this mechanism, the proximal His residues on only the α subunits, only the β subunits, and both the α and β subunits, were mutagenically changed to Gly and the missing imidazole ring of the proximal His was replaced by imidazole (which a variety of evidence indicates ligands the heme Fe as does the proximal His). This, in effect, detaches the proximal His from the protein, thereby cutting the covalent link that presumably connects the ligand-induced movement of the Fe into the heme plane to the accompanying movement of helix F. This proximal detachment is observed to significantly increase hemoglobin's ligand-binding affinity, to reduce its cooperativity, and to prevent its quaternary switching from the T state to the R state, in accord with the Perutz model. However, these mutant hemoglobins exhibit a small amount of residual cooperativity, suggesting that the heme groups also communicate via pathways that do not require covalent coupling between the F helix and the proximal His. These pathways may involve movements of the distal His residues of the α and β subunits and/or the movement of Val E11 of the β subunits, all of whose sidechains must move aside when ligand binds to Hb.

Barrick, D., Ho, N.T., Simplaceanu, V., Dahlquist, F.W., and Ho, C., A test of the role of the proximal histidines in the Perutz model for cooperativity in haemoglobin, *Nature Struct. Biol.* **4,** 78–83 (1997).

(d) Additional References

Hardison, R.C., A brief history of hemoglobins: Plant, animal, protist, and bacteria, *Proc. Natl. Acad. Sci.* **93,** 5675–5679 (1996).

3. ABNORMAL HEMOGLOBINS

Cao, Z. and Ferrone, F.A., A 50th order reaction predicted and observed for sickle hemoglobin nucleation, *J. Mol. Biol.* **256,** 219–222 (1996).

Sanders, K.E., Ackers, G., and Sligar, S., Engineering and design of blood substitutes, *Curr. Opin. Struct. Biol.* **6,** 534–540 (1996).

Chapter 10

SUGARS AND POLYSACCHARIDES

Ball, S, Guan, H.-P., James, M., Myers, A., Keeling, P., Mouille, G., Buléon, A., Colonna, P., and Priess, J., From glycogen to amylopectin: A model for the biogenesis of the plant starch granule, *Cell* **86,** 349–351 (1996).