

2

EDITION

FUNDAMENTALS OF

BIOCHEMISTRY

LIFE AT THE MOLECULAR LEVEL

DONALD VOET

University of Pennsylvania

JUDITH G. VOET

Swarthmore College

CHARLOTTE W. PRATT

Seattle, Washington



WILEY

JOHN WILEY & SONS, INC.

Brigham and Women's Hospital - EXHIBIT 2025

PreOmics GmbH and Biognosys AG v. The Brigham and

Women's Hospital, Inc.

IPR2024-01473

Senior Acquisitions Editor: Patrick Fitzgerald
Senior Developmental Editor: Ellen Ford
Marketing Manager: Amanda Wygal
Cover/Text Designer: Madelyn Lesure
Production Editor: Sandra Dumas
Media Editor: Thomas Kulesa
Photo Editor: Hilary Newman
Photo Researcher: Elyse Rieder
Illustration Editor: Sigmund Malinowski
Production Management Services: Suzanne Ingrao/Ingrao Associates
Cover Photos:

(Left photo): Geoffrey Gove/The Image Bank/Getty Images.

(Right photos, top to bottom): DNA polymerase courtesy of Gabriel Waksman, Washington University School of Medicine; GlnRS-tRNA^{Gln} based on an X-ray structure by Thomas Steitz, Yale University; RNA polymerase II based on an X-ray structure by Roger Kornberg, Stanford University School of Medicine; large ribosomal subunit courtesy of Harry Noller, University of California at Santa Cruz; DNA © Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute; rights owned by HHMI. Reproduction by permission only.

This book was typeset in 10/12 Times Ten with Trade Gothic at GTS Companies and printed and bound by Von Hoffmann Corporation. The cover was printed by Lehigh Press.

The paper in this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

This book is printed on acid-free paper. ©

Copyright © 2006 by Donald Voet, Judith G. Voet, and Charlotte W. Pratt. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008. To order books or for customer service call 1-800-CALL-WILEY(225-5945).

Voet, Donald; Voet, Judith G.; Pratt, Charlotte W.
Fundamentals of Biochemistry, Second Edition

ISBN 0-471-21495-7

Printed in the United States of America.

10 9 8 7 6 5 4 3 2

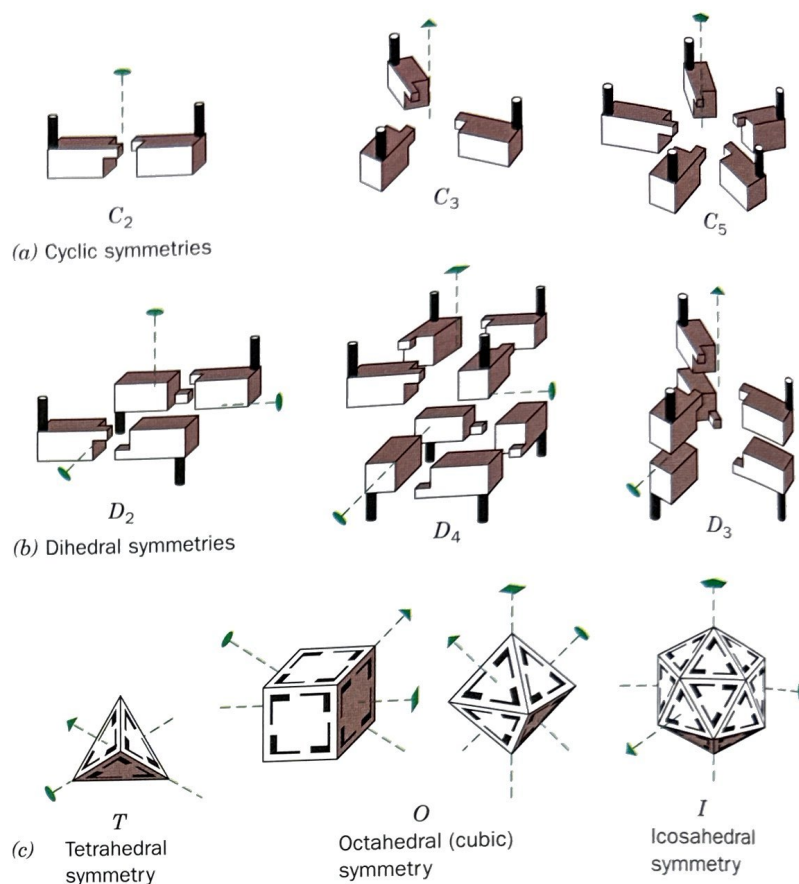


Figure 6-34 Some symmetries of oligomeric proteins. The oval, the triangle, the square, and the pentagon at the ends of the dashed green lines indicate, respectively, the unique twofold, threefold, fourfold, and fivefold rotational axes of the objects shown. (a) Assemblies with cyclic (C) symmetry. (b) Assemblies with dihedral (D) symmetry. In these objects, a 2-fold axis is perpendicular to another rotational axis. (c) Assemblies with the rotational symmetries of a tetrahedron (T), a cube or octahedron (O), and an icosahedron (I). [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] See the Animated Figures.

In the simplest type of rotational symmetry, **cyclic symmetry**, protomers are related by a single axis of rotation (Fig. 6-34a). Objects with two-, three-, or n -fold rotational axes are said to have C_2 , C_3 , or C_n symmetry, respectively. C_2 symmetry is the most common; higher cyclic symmetries are relatively rare.

Dihedral symmetry (D_n), a more complicated type of rotational symmetry, is generated when an n -fold rotation axis intersects a twofold rotation axis at right angles (Fig. 6-34b). An oligomer with D_n symmetry consists of $2n$ protomers. D_2 symmetry is the most common type of dihedral symmetry in proteins.

Other possible types of rotational symmetry are those of a tetrahedron, cube, and icosahedron (Fig. 6-34c). Some multienzyme complexes and spherical viruses are built on these geometric plans.

4 Protein Stability

Incredible as it may seem, thermodynamic measurements indicate that *native proteins are only marginally stable under physiological conditions*. The free energy required to denature them is $\sim 0.4 \text{ kJ} \cdot \text{mol}^{-1}$ per amine acid residue, so a fully folded 100-residue protein is only about $40 \text{ kJ} \cdot \text{mol}^{-1}$ more stable than its unfolded form (for comparison, the energy required to break a typical hydrogen bond is $\sim 20 \text{ kJ} \cdot \text{mol}^{-1}$). The various noncovalent influences on proteins—hydrophobic effects, electrostatic interactions, and hydrogen bonding—each have energies that may total thousands of kilojoules per mole over an entire protein molecule. Consequently, a pro-

Protein structure is the result of a delicate balance among powerful counterbalancing forces. In this section, we discuss the forces that stabilize proteins and the processes by which proteins achieve their most stable folded state.

A Forces That Stabilize Protein Structure

Protein structures are governed primarily by hydrophobic effects and, to a lesser extent, by interactions between polar residues, and by other types of bonds.

The Hydrophobic Effect. The hydrophobic effect, which causes nonpolar substances to minimize their contacts with water (Section 2-1C), is the major determinant of native protein structure. The aggregation of nonpolar side chains in the interior of a protein is favored by the increase in entropy of the water molecules that would otherwise form ordered “cages” around the hydrophobic groups. The combined hydrophobic and hydrophilic tendencies of individual amino acid residues in proteins can be expressed as **hydropathies** (Table 6-2). The greater a side chain’s hydropathy, the more likely it is to occupy the interior of a protein and vice versa. Hydropathies are good predictors of which portions of a polypeptide chain are inside a protein, out of contact with the aqueous solvent, and which portions are outside (Fig. 6-35).

Site-directed mutagenesis experiments in which individual interior residues have been replaced by a number of others suggest that the factors that affect stability are, in order, the hydrophobicity of the substituted residue, its steric compatibility, and, last, the volume of its side chain.

Electrostatic Interactions. In the closely packed interiors of native proteins, van der Waals forces, which are relatively weak (Section 2-1A), are nevertheless an important stabilizing influence. This is because these forces act over only short distances and hence are lost when the protein is unfolded.

Perhaps surprisingly, *hydrogen bonds, which are central features of protein structures, make only minor contributions to protein stability.* This is because hydrogen-bonding groups in an unfolded protein form hydrogen bonds with water molecules. Thus the contribution of a hydrogen bond to the stability of a native protein is the small difference in hydrogen bonding free energies between the native and unfolded states (-2 to $8 \text{ kJ}\cdot\text{mol}^{-1}$ as determined by site-directed mutagenesis studies). Nevertheless, hydrogen bonds are important determinants of native protein structures, because if a protein folded in a way that prevented a hydrogen bond from forming, the stabilizing energy of that hydrogen bond would be lost. Hydrogen bonding therefore fine-tunes tertiary structure by “selecting” the unique native structure of a protein from among a relatively small number of hydrophobically stabilized conformations.

Table 6-2 Hydropathy Scale for Amino Acid Side Chains

Side Chain	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

Source: Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* **157**, 110 (1982).

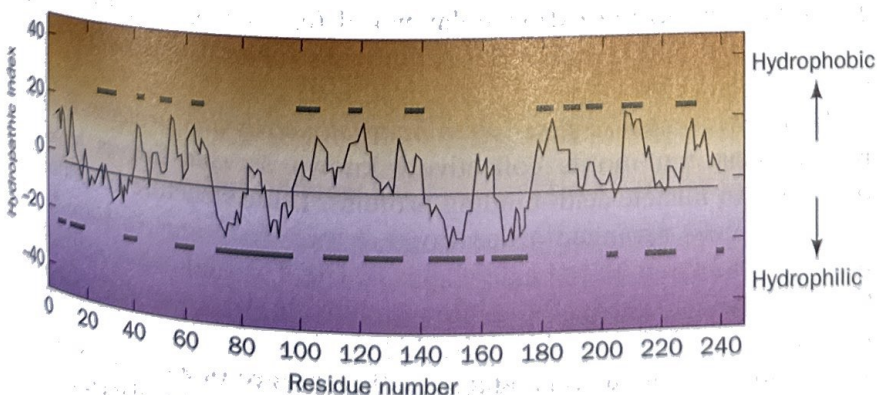


Figure 6-35 A hydropathic index plot for bovine chymotrypsinogen. The sum of the hydropathies of nine consecutive residues is plotted versus residue sequence number. A large positive hydropathic index indicates a hydrophobic region of the polypeptide, whereas a large negative value indicates a hydrophilic region. The upper bars denote the protein’s interior regions, as determined by X-ray crystallography, and the lower bars denote the protein’s exterior regions. [After Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* **157**, 111 (1982).]

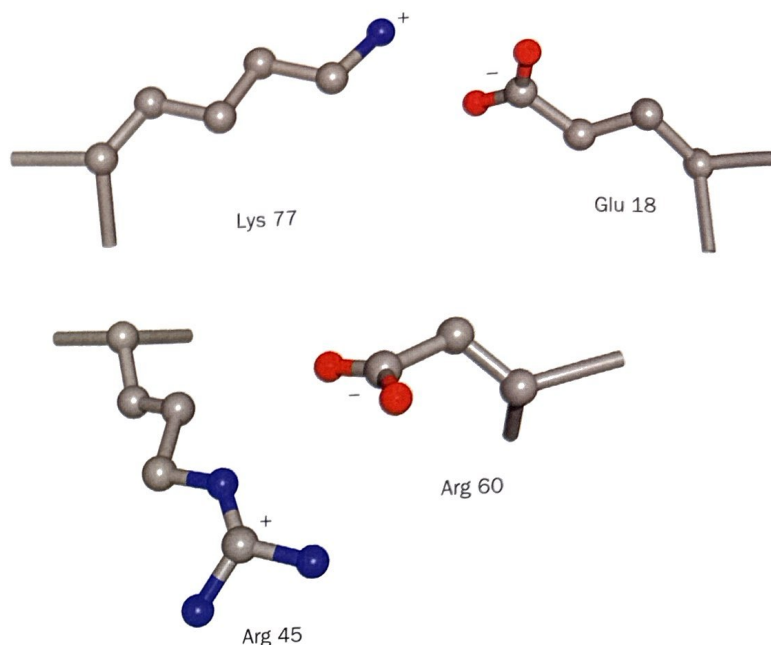


Figure 6-36 Examples of ion pairs in myoglobin. In each case, oppositely charged side chain groups from residues far apart in sequence closely approach each other through the formation of ion pairs.

The association of two ionic protein groups of opposite charge (e.g., Lys and Asp) is known as an **ion pair** or **salt bridge**. About 75% of the charged residues in proteins are members of ion pairs that are located mostly on the protein surface (Fig. 6-36). Despite the strong electrostatic attraction between the oppositely charged members of an ion pair, these interactions contribute little to the stability of a native protein. This is because the free energy of an ion pair's charge-charge interactions usually fails to compensate for the loss of entropy of the side chains and the loss of solvation free energy when the charged groups form an ion pair. This accounts for the observation that ion pairs are poorly conserved among homologous proteins.

Chemical Cross-Links. Disulfide bonds (Fig. 4-6) within and between polypeptide chains form as a protein folds to its native conformation. Some polypeptides whose Cys residues have been derivatized or mutagenically replaced to prevent disulfide bond formation can still assume their fully active conformations, suggesting that disulfide bonds are not essential stabilizing forces. They may, however, be important for “locking in” a particular backbone folding pattern as the protein proceeds from its fully extended state to its mature form.

Disulfide bonds are rare in intracellular proteins because the cytoplasm is a reducing environment. Most disulfide bonds occur in proteins that are secreted from the cell into the more oxidizing extracellular environment. The relatively hostile extracellular world (e.g., uncontrolled temperature and pH) apparently requires the additional structural constraints conferred by disulfide bonds.

Metal ions may also function to internally cross-link proteins. For example, at least ten motifs collectively known as **zinc fingers** have been described in nucleic acid-binding proteins. These structures contain about 25–60 residues arranged around one or two Zn^{2+} ions that are tetrahedrally coordinated by the side chains of Cys, His, and occasionally Asp or Glu (Fig. 6-37). The Zn^{2+} ion allows relatively short stretches of polypeptide chain to fold into stable units that can interact with nucleic acids. Zinc fingers are too small to be stable in the absence of Zn^{2+} . Zinc is ideally suited to its structural role in intracellular proteins: Its filled *d* electron shell

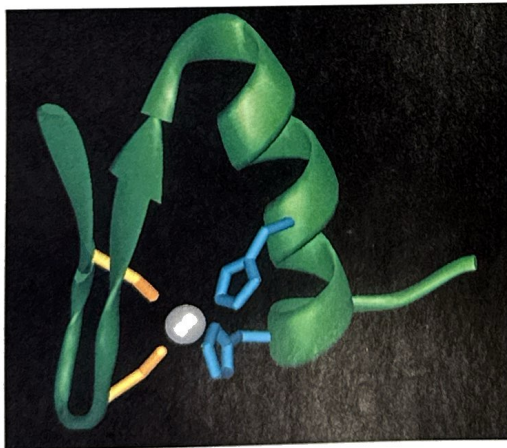


Figure 6-37 A zinc finger motif. This structure, from the DNA-binding protein Zif268, is known as a $\text{Cys}_2\text{-His}_2$ zinc finger because the zinc atom (silver sphere) is coordinated by two Cys residues (yellow) and two His residues (cyan). [Based on an X-ray structure by Carl Pabo, MIT. PDBid 1ZAA.]

permits it to interact strongly with a variety of ligands (e.g., sulfur, nitrogen, or oxygen) from different amino acid residues. In addition, zinc has only one stable oxidation state (unlike, for example, copper and iron), so it does not undergo oxidation–reduction reactions in the cell.

B Protein Dynamics

The plethora of forces acting to stabilize proteins as well as the static way that their structures are usually portrayed may leave the false impression that proteins have fixed and rigid structures. In fact, *proteins are flexible and rapidly fluctuating molecules whose structural mobilities are functionally significant*. Groups ranging in size from individual side chains to entire domains or subunits may be displaced by up to several angstroms through random intramolecular movements or in response to a trigger such as the binding of a small molecule. Extended side chains, such as Lys, and the N- and C-termini of polypeptide chains are especially prone to wave around in solution because there are few forces holding them in place.

Theoretical calculations by Martin Karplus indicate that a protein's native structure probably consists of a large collection of rapidly interconverting conformations that have essentially equal stabilities (Fig. 6-38). Conformational flexibility, or **breathing**, with structural displacement of up to ~ 2 Å, allows small molecules to diffuse in and out of the interior of certain proteins.

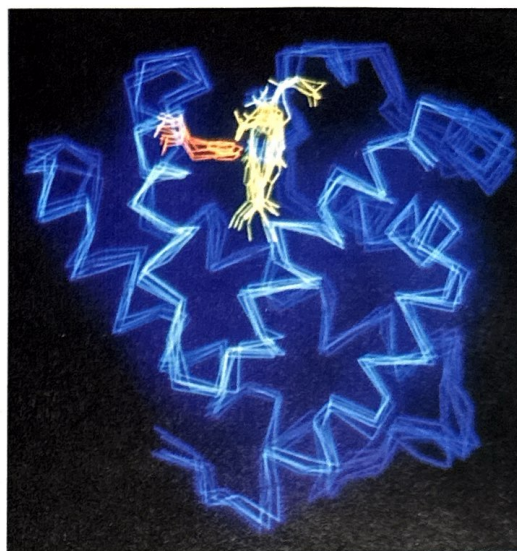


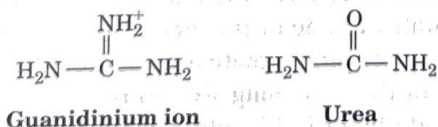
Figure 6-38 Molecular dynamics of myoglobin.

Several “snapshots” of the protein calculated at intervals of 5×10^{-12} s are superimposed. The backbone is blue, the heme group is yellow, and the His side chain linking the heme to the protein is orange. [Courtesy of Martin Karplus, Harvard University.]

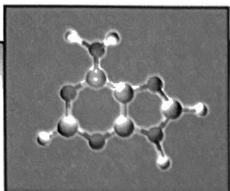
C Protein Denaturation and Renaturation

The low conformational stabilities of native proteins make them easily susceptible to denaturation by altering the balance of the weak nonbonding forces that maintain the native conformation. Proteins can be denatured by a variety of conditions and substances:

1. Heating causes a protein's conformationally sensitive properties, such as optical rotation (Section 4-2), viscosity, and UV absorption, to change abruptly over a narrow temperature range. Such a sharp transition indicates that the entire polypeptide unfolds or “melts” **cooperatively**, that is, nearly simultaneously. Most proteins have melting temperatures that are well below 100°C . Among the exceptions are the proteins of thermophilic bacteria (Box 6-3).
2. pH variations alter the ionization states of amino acid side chains, thereby changing protein charge distributions and hydrogen-bonding requirements.
3. Detergents associate with the nonpolar residues of a protein, thereby interfering with the hydrophobic interactions responsible for the protein's native structure.
4. The **chaotropic agents** guanidinium ion and urea,



in concentrations in the range 5 to 10 M, are the most commonly used protein denaturants. Chaotropic agents are ions or small organic molecules that increase the solubility of nonpolar substances in water. Their effectiveness as denaturants stems from their ability to disrupt hydrophobic interactions, although their mechanism of action is not well understood.



BOX 6 - 3

*Perspectives in Biochemistry**Thermostable Proteins*

Certain species of bacteria known as **hyperthermophiles** grow at temperatures near 100°C. They live in such places as hot springs and submarine hydrothermal vents, with the most extreme, the archeon *Pyrolobus fumarii*, able to grow at temperatures as high as 113°C. These organisms have many of the same metabolic pathways as do **mesophiles** (organisms that grow at “normal” temperatures). Yet most mesophilic proteins denature at temperatures where hyperthermophiles thrive. What is the structural basis for the thermostability of hyperthermophilic proteins?

The difference in the thermal stabilities of the corresponding (hyper)thermophilic and mesophilic proteins does not exceed $\sim 100 \text{ kJ} \cdot \text{mol}^{-1}$, the equivalent of a few noncovalent interactions. This is probably why comparisons of the X-ray structures of hyperthermophilic enzymes with their mesophilic counterparts have failed to reveal any striking differences between them. These proteins exhibit some variations in secondary structure but no more than would be expected for homologous proteins from distantly related mesophiles. However, several of these thermostable enzymes have a superabundance of salt bridges on their surfaces, many of which are arranged in extensive networks containing up to 18 side chains.

The idea that salt bridges can stabilize a protein structure appears to contradict the conclusion of Section 6-4A that ion pairs are, at best, marginally stable. The key to this apparent

paradox is that *the salt bridges in thermostable proteins form networks*. Thus, the gain in charge–charge free energy on associating a third charged group with an ion pair is comparable to that between the members of this ion pair, whereas the free energy lost on desolvating and immobilizing the third side chain is only about half that lost in bringing together the first two side chains. The same, of course, is true for the addition of a fourth, fifth, etc., side chain to a salt bridge network.

Not all thermostable proteins have such a high incidence of salt bridges. Structural comparisons suggest that these proteins are stabilized by a combination of small effects, the most important of which are an increased size of the protein's hydrophobic core, an increased size in the interface between its domains and/or subunits, and a more tightly packed core as evidenced by a reduced surface-to-volume ratio.

The fact that the proteins of hyperthermophiles and mesophiles are homologous and carry out much the same functions indicates that mesophilic proteins are by no means maximally stable. This, in turn, strongly suggests *that the marginal stability of most proteins under physiological conditions (averaging $\sim 0.4 \text{ kJ} \cdot \text{mol}^{-1}$ of amino acid residues) is an essential property that has arisen through evolutionary design*. Perhaps this marginal stability helps confer the structural flexibility that many proteins require to carry out their physiological functions.

Denatured Proteins Can Be Renatured. In 1957, the elegant experiments of Christian Anfinsen on **ribonuclease A (RNase A)** showed that proteins can be denatured reversibly. RNase A, a 124-residue single-chain protein, is completely unfolded and its four disulfide bonds reductively cleaved in an 8 M urea solution containing 2-mercaptoethanol. Dialyzing away the urea and reductant and exposing the resulting solution to O_2 at pH 8 (which oxidizes the SH groups to form disulfides) yields a protein that is virtually 100% enzymatically active and physically indistinguishable from native RNase A (Fig. 6-39). The protein must therefore **renature** spontaneously.

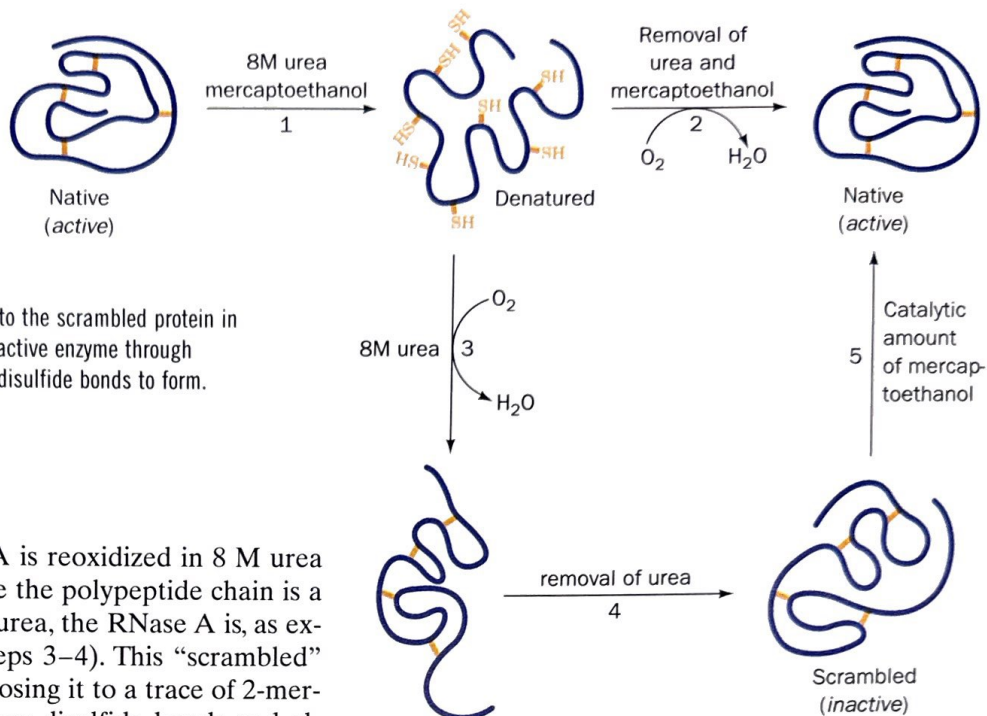
The renaturation of RNase A demands that its four disulfide bonds re-form. The probability of one of the eight Cys residues randomly forming a disulfide bond with its proper mate among the other seven Cys residues is $1/7$; that of one of the remaining six Cys residues then randomly forming its proper disulfide bond is $1/5$; etc. The overall probability of RNase A re-forming its four native disulfide links at random is

$$\frac{1}{7} \times \frac{1}{5} \times \frac{1}{3} \times \frac{1}{1} = \frac{1}{105}$$

Clearly, the disulfide bonds do not randomly re-form under renaturing conditions, since, if they did, only 1% of the refolded protein would be cat-

Figure 6-39 Denaturation and renaturation of RNase A.

The polypeptide is represented by a purple line, with its disulfide bonds in yellow. After the protein has been denatured and its disulfide bonds cleaved (1), it will renature in the presence of O_2 when the denaturant (urea) and reductant (mercaptoethanol) are removed (2). If the disulfide bonds are allowed to re-form (3) before the urea is removed (4), the bonds form at random and the resulting protein is enzymatically inactive. Adding a small amount of mercaptoethanol to the scrambled protein in the absence of O_2 (5) catalyzes its conversion to the active enzyme through disulfide interchange reactions that allow the native disulfide bonds to form.



alytically active. Indeed, if the RNase A is reoxidized in 8 M urea so that its disulfide bonds re-form while the polypeptide chain is a random coil, then after removal of the urea, the RNase A is, as expected, only $\sim 1\%$ active (Fig. 6-39, Steps 3–4). This “scrambled” protein can be made fully active by exposing it to a trace of 2-mercaptoethanol, which breaks the improper disulfide bonds and allows the proper bonds to form. *Anfinsen’s work demonstrated that proteins can fold spontaneously into their native conformations under physiological conditions. This implies that a protein’s primary structure dictates its three-dimensional structure.*

5 Protein Folding

Studies of protein stability and renaturation suggest that protein folding is directed largely by the residues that occupy the interior of the folded protein. But *how* does a protein fold to its native conformation? One might guess that this process occurs through the protein’s random exploration of all the conformations available to it until it eventually stumbles onto the correct one. A simple calculation first made by Cyrus Levinthal, however, convincingly demonstrates that this cannot possibly be the case: Assume that an n -residue protein’s 2^n torsion angles, ϕ and ψ , each have three stable conformations. This yields $3^{2n} \approx 10^n$ possible conformations for the protein (a gross underestimate because we have completely neglected its side chains). Then, if the protein could explore a new conformation every 10^{-13} s (the rate at which single bonds reorient), the time t , in seconds, required for the protein to explore all the conformations available to it is

$$t = \frac{10^n}{10^{13}}$$

For a small protein of 100 residues, $t = 10^{87}$ s, which is immensely greater than the apparent age of the universe (20 billion years, or 6×10^{17} s). Clearly, proteins must fold more rapidly than this.

A Protein Folding Pathways

Experiments have shown that many proteins fold to their native conformations in less than a few seconds. This is because *proteins fold to their native conformations via directed pathways rather than stumbling on them through random conformational searches*. Thus, as a protein folds, its confor-