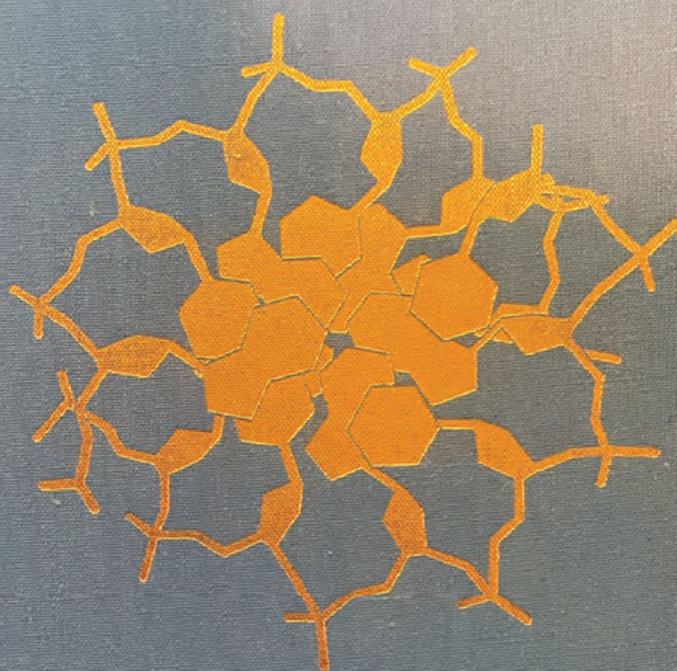


SIXTH EDITION

Biochemistry



Jeremy M. Berg
John L. Tymoczko
Lubert Stryer

Publisher: Sara Tenney
Senior Acquisitions Editor: Kate Ahr
Marketing Managers: Sarah Martin, John Britch
Senior Developmental Editor: Susan Moran
Media Editor: Alysia Baker
Supplements Editors: Nick Tymoczko, Deena Goldman
Photo Editor: Bianca Moscatelli
Design Manager: Diana Blume
Text Designer: Patrice Sheridan
Senior Project Editor: Georgia Lee Hadler
Manuscript Editor: Patricia Zimmerman
Illustrations: Jeremy Berg with Network Graphics
Senior Illustration Coordinator: Bill Page
Production Coordinator: Susan Wein
Composition: Techbooks
Printing and Binding: RR Donnelley

Library of Congress Cataloging-in-Publication Data

Berg, Jeremy Mark.

Biochemistry / Jeremy M. Berg, John L. Tymoczko, Lubert Stryer.—6th ed.
p. cm.

Includes bibliographical references and index.

ISBN 0-7167-8724-5 hardcover

1. Biochemistry. I. Tymoczko, John L., II. Stryer, Lubert III. Title.

QP514.2.S66 2006

572—dc22

2005052751

ISBN-13: 978-0-7167-8724-2

ISBN-10: 0-7167-8724-5

©2007, 2002 by W. H. Freeman and Company;

© 1975, 1981, 1988, 1995 by Lubert Stryer

All rights reserved

Printed in the United States of America

Third printing

W. H. Freeman and Company

41 Madison Avenue

New York, NY 10010

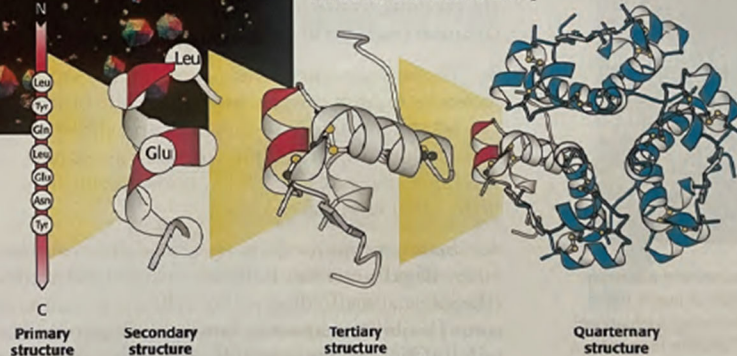
Houndmills, Basingstoke RG21 6XS, England

www.whfreeman.com

Protein Composition and Structure



Crystals of human insulin. Insulin is a protein hormone, crucial for maintaining blood sugar at appropriate levels. (Below) Chains of amino acids in a specific sequence (the primary structure) define a protein such as insulin. These chains fold into well-defined structures (the tertiary structure)—in this case, a single insulin molecule. Such structures assemble with other chains to form arrays such as the complex of six insulin molecules shown at the far right (the quaternary structure). These arrays can often be induced to form well-defined crystals (photograph at left), which allows determination of these structures in detail. [Photograph from Alfred Pasieka/Peter Arnold.]



Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. They function as catalysts, transport and store other molecules such as oxygen, provide mechanical support and immune protection, generate movement, transmit nerve impulses, and control growth and differentiation. Indeed, much of this book will focus on understanding what proteins do and how they perform these functions.

Several key properties enable proteins to participate in a wide range of functions.

1. *Proteins are linear polymers built of monomer units called amino acids*, which are linked end to end. Remarkably, proteins spontaneously fold up into three-dimensional structures that are determined by the sequence of amino acids in the protein polymer. Protein function is directly dependent on this three-dimensional structure (Figure 2.1). Thus, *proteins are the embodiment of the transition from the one-dimensional world of sequences to the three-dimensional world of molecules capable of diverse activities*.

2. *Proteins contain a wide range of functional groups*. These functional groups include alcohols, thiols, thioethers, carboxylic acids, carboxamides, and a variety of basic groups. Most of these groups are chemically reactive. When

Outline

- 2.1 Proteins Are Built from a Repertoire of 20 Amino Acids
- 2.2 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains
- 2.3 Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops
- 2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores
- 2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures
- 2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

Figure 2.1 Structure dictates function. A protein component of the DNA replication machinery surrounds a section of DNA double helix depicted as a cylinder. The protein, which consists of two identical subunits (shown in red and yellow), acts as a clamp that allows large segments of DNA to be copied without the replication machinery dissociating from the DNA. [Drawn from 2POL.pdb.]

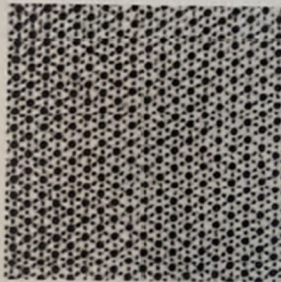
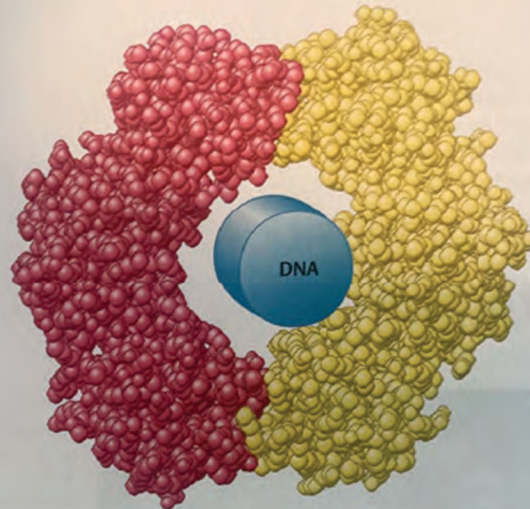


Figure 2.2 A complex protein assembly. An electron micrograph of insect flight tissue in cross section shows a hexagonal array of two kinds of protein filaments. [Courtesy of Dr. Michael Reedy.]

combined in various sequences, this array of functional groups accounts for the broad spectrum of protein function. For instance, their reactive properties are essential to the function of *enzymes*, the proteins that catalyze specific chemical reactions in biological systems (see Chapters 8 through 10).

3. *Proteins can interact with one another and with other biological macromolecules to form complex assemblies.* The proteins within these assemblies can act synergistically to generate capabilities that individual proteins may lack (Figure 2.2). Examples of these assemblies include macromolecular machines that replicate DNA, transmit signals within cells, and carry out many other essential processes.

4. *Some proteins are quite rigid, whereas others display a considerable flexibility.* Rigid units can function as structural elements in the cytoskeleton (the internal scaffolding within cells) or in connective tissue. Proteins with some flexibility may act as hinges, springs, or levers that are crucial to protein function, to the assembly of proteins with one another and with other molecules into complex units, and to the transmission of information within and between cells (Figure 2.3).

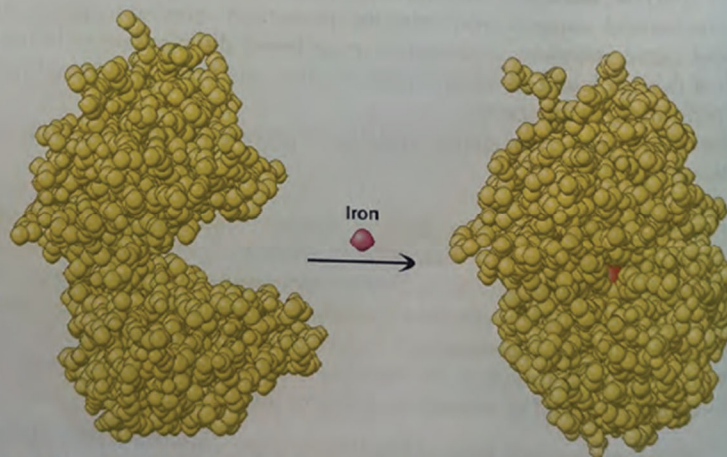
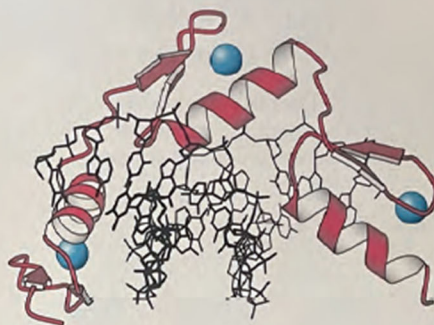


Figure 2.3 Flexibility and function. On binding iron, the protein lactoferrin undergoes a substantial change in conformation that allows other molecules to distinguish between the iron-free and the iron-bound forms. [Drawn from 1LFH.pdb and 1LFG.pdb.]

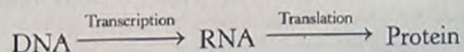
DNA, RNA, and the Flow of Genetic Information



Having genes in common accounts for the resemblance of a mother to her daughters. Genes must be expressed to exert an effect, and proteins regulate such expression. One such regulatory protein, a zinc-finger protein (zinc ion is blue, protein is red), is shown bound to a control region of DNA (black). [(Left) Barnaby Hall/Photonica. (Right) Drawn from 1AAY.pdb.]

DNA and RNA are long linear polymers, called nucleic acids, that carry information in a form that can be passed from one generation to the next. These macromolecules consist of a large number of linked nucleotides, each composed of a sugar, a phosphate, and a base. Sugars linked by phosphates form a common backbone that plays a structural role, whereas the *sequence of bases along a nucleic acid chain carries genetic information*. The DNA molecule has the form of a *double helix*, a helical structure consisting of two complementary nucleic acid strands. *Each strand serves as the template for the other in DNA replication*. The genes of all cells and many viruses are made of DNA. Some viruses, however, use RNA as their genetic material.

Genes specify the kinds of proteins that are made by cells, but DNA is not the direct template for protein synthesis. Rather, a DNA strand is copied into a class of RNA molecules called *messenger RNA (mRNA)*, which are the *information-carrying intermediates* in protein synthesis. This process of *transcription* is followed by *translation*, the synthesis of proteins according to instructions given by mRNA templates. Thus, the flow of genetic information, or *gene expression*, in normal cells is



Outline

- 4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar-Phosphate Backbone
- 4.2 A Pair of Nucleic Acid Chains with Complementary Sequences Can Form a Double-Helical Structure
- 4.3 DNA Is Replicated by Polymerases That Take Instructions from Templates
- 4.4 Gene Expression Is the Transformation of DNA Information into Functional Molecules
- 4.5 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point
- 4.6 Most Eukaryotic Genes Are Mosaics of Introns and Exons

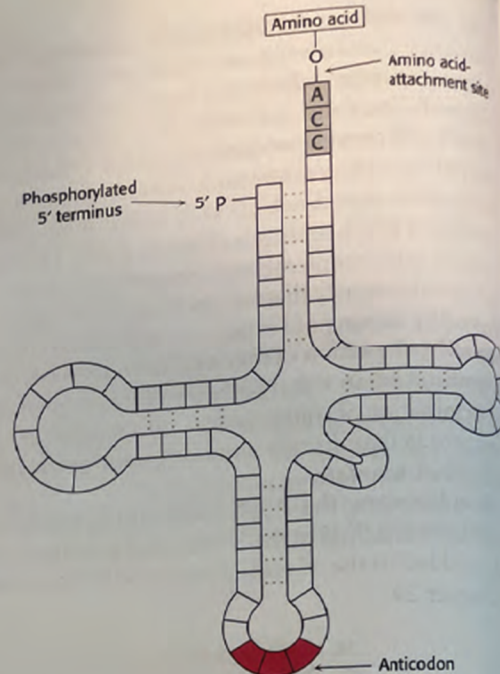


Figure 4.31 General structure of an aminoacyl-tRNA. The amino acid is attached at the 3' end of the RNA. The anticodon is the template-recognition site. Notice that the tRNA has a cloverleaf structure with many hydrogen bonds (green dots) between bases.

4.5 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point

The *genetic code* is the relation between the sequence of bases in DNA (or in RNA transcripts) and the sequence of amino acids in proteins. Experiments by Marshall Nirenberg, Har Gobind Khorana, Francis Crick, Sydney Brenner and others established the following features of the genetic code by 1961:

1. *Three nucleotides encode an amino acid.* Proteins are built from a basic set of 20 amino acids, but there are only four bases. Simple calculations show that a minimum of three bases is required to encode at least 20 amino acids. Genetic experiments showed that *an amino acid is in fact encoded by a group of three bases, or codon.*
2. *The code is nonoverlapping.* Consider a base sequence ABCDEF. In an overlapping code, ABC specifies the first amino acid, BCD the next, CDE the next, and so on. In a nonoverlapping code, ABC designates the first amino acid, DEF the second, and so forth. Genetic experiments again established the code to be nonoverlapping.
3. *The code has no punctuation.* In principle, one base (denoted as Q might serve as a "comma" between groups of three bases.

... QABCQDEFQGHIOJKI.Q ...

This is not the case. Rather, the sequence of bases is read sequentially from a fixed starting point, without punctuation.

4. *The genetic code is degenerate.* Most amino acids are encoded by more than one codon. There are 64 possible base triplets and only 20 amino acids and in fact 61 of the 64 possible triplets specify particular amino acids. Three triplets (called *stop codons*) designate the termination of translation. Thus, for most amino acids, there is more than one code word.

Major Features of the Genetic Code

All 64 codons have been deciphered (Table 4.4). Because the code is highly degenerate, only tryptophan and methionine are encoded by just one triplet each. Each of the other 18 amino acids is encoded by two or more. Indeed, leucine, arginine, and serine are specified by six codons each. The number of codons for a particular amino acid correlates with its frequency of occurrence in proteins.

125

4.5 The Genetic Code

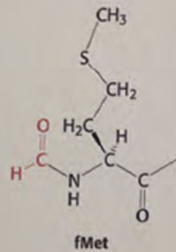
TABLE 4.4 The genetic code

First position (5' end)	Second Position				Third position (3' end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Note: This table identifies the amino acid encoded by each triplet. For example, the codon 5' AUG 3' on mRNA specifies methionine, whereas CAU specifies histidine. UAA, UAG, and UGA are termination signals. AUG is part of the initiation signal, in addition to coding for internal methionine residues.

Codons that specify the same amino acid are called *synonyms*. For example, CAU and CAC are synonyms for histidine. Note that synonyms are not distributed haphazardly throughout the genetic code (depicted in Table 4.4). In the table, an amino acid specified by two or more synonyms occupies a single box (unless it is specified by more than four synonyms). The amino acids in a box are specified by codons that have the same first two bases but differ in the third base, as exemplified by GUU, GUC, GUA, and GUG. Thus, *most synonyms differ only in the last base of the triplet*. Inspection of the code shows that XYC and XYU always encode the same amino acid, whereas XYG and XYA usually encode the same amino acid. The structural basis for these equivalences of codons will become evident when we consider the nature of the anticodons of tRNA molecules (Section 30.3).

What is the biological significance of the extensive degeneracy of the genetic code? If the code were not degenerate, 20 codons would designate amino acids and 44 would lead to chain termination. The probability of mutating to chain termination would therefore be much higher with a non-degenerate code. Chain-termination mutations usually lead to inactive proteins, whereas substitutions of one amino acid for another are usually rather harmless. Moreover, the code is constructed such that a change to any single nucleotide base of a codon results in a synonym or an amino acid with similar chemical properties. Thus, *degeneracy minimizes the deleterious effects of mutations*.



Messenger RNA Contains Start and Stop Signals for Protein Synthesis

Messenger RNA is translated into proteins on *ribosomes*, large molecular complexes assembled from proteins and ribosomal RNA. How is mRNA interpreted by the translation apparatus? The start signal for protein synthesis is complex in bacteria. Polypeptide chains in bacteria start with a modified amino acid—namely, **formylmethionine (fMet)**. A special tRNA, the initiator tRNA, carries fMet. This fMet-tRNA recognizes the codon AUG or, less frequently, GUG. However, AUG is also the codon for an internal methionine residue, and GUG is the codon for an internal valine residue. Hence, the signal for the first amino acid in a prokaryotic polypeptide chain must be more complex than that for all subsequent ones. AUG (or GUG) is only part of the initiation signal (Figure 4.32). In bacteria, the initiating AUG (or GUG) codon is preceded several nucleotides away by a purine-rich sequence, called the *Shine-Dalgarno sequence*, that base-pairs with a complementary sequence in a ribosomal RNA molecule (Section 30.2). In eukaryotes, the AUG closest to the 5' end of an mRNA molecule is usually the start signal for protein synthesis. This particular AUG is read by an initiator tRNA conjugated to methionine. After the initiator AUG has been located, the *reading frame* is established—groups of three nonoverlapping nucleotides are defined, beginning with the initiator AUG codon.

As already mentioned, UAA, UAG, and UGA designate *chain termination*. These codons are read not by tRNA molecules but rather by specific proteins called *release factors* (Section 30.3). Binding of the release factors to the ribosomes releases the newly synthesized protein.

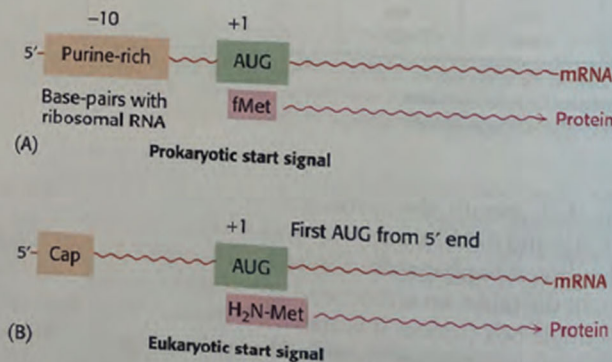


Figure 4.32 Initiation of protein synthesis. Start signals are required for the initiation of protein synthesis in (A) prokaryotes and (B) eukaryotes.

The Genetic Code Is Nearly Universal

Is the genetic code the same in all organisms? The base sequences of many wild-type and mutant genes are known, as are the amino acid sequences of their encoded proteins. For each mutant, the nucleotide change in the gene and the amino acid change in the protein are as predicted by the genetic code. Furthermore, mRNAs can be correctly translated by the protein-synthesizing machinery of very different species. For example, human hemoglobin mRNA is correctly translated by a wheat-germ extract, and bacteria efficiently express recombinant DNA molecules encoding human proteins such as insulin. These experimental findings strongly suggest that the genetic code is universal.

A surprise was encountered when the sequence of human mitochondrial DNA became known. Human mitochondria read UGA as a codon for tryptophan rather than as a stop signal (Table 4.5). Furthermore, AGA and AGG are read as stop signals rather than as codons for arginine, and AU