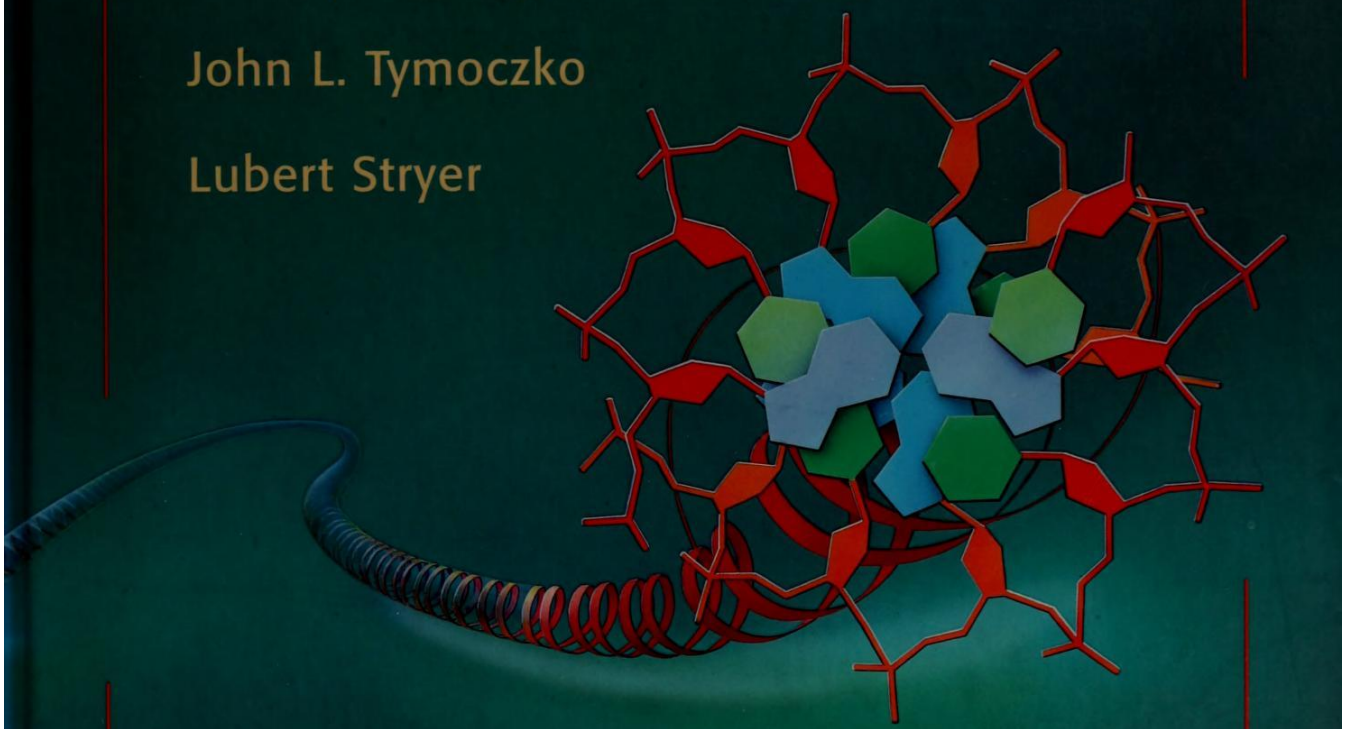
BIOCHEMISTRY

2001

Jeremy M. Berg

John L. Tymoczko

Lubert Stryer



FIFTH EDITION

BIOCHEMISTRY

FIFTH EDITION

Jeremy M. Berg
Johns Hopkins University School of Medicine

John L. Tymoczko
Carleton College

Lubert Stryer
Stanford University

Web content by
Neil D. Clarke
Johns Hopkins University School of Medicine

W. H. Freeman and Company
New York

PUBLISHER: **Michelle Julet**

DEVELOPMENT EDITOR: **Susan Moran**

NEW MEDIA DEVELOPMENT EDITOR: **Sonia Divittorio**

NEW MEDIA AND SUPPLEMENTS EDITOR: **Mark Santee**

MEDIA DEVELOPERS: **CADRE design; molvisions.com—3D molecular visualization**

MARKETING MANAGER: **Carol Coffey**

PROJECT EDITOR: **Georgia Lee Hadler**

MANUSCRIPT EDITOR: **Patricia Zimmerman**

COVER AND TEXT DESIGN: **Victoria Tomaselli**

COVER ILLUSTRATION: **Tomo Narashima**

ILLUSTRATION COORDINATOR: **Cecilia Varas**

ILLUSTRATIONS: **Jeremy Berg with Network Graphics**

PHOTO EDITOR: **Vikii Wong**

PHOTO RESEARCHER: **Dena Betz**

PRODUCTION COORDINATOR: **Julia DeRosa**

COMPOSITION: **TechBooks**

MANUFACTURING: **RR Donnelley & Sons Company**

Library of Congress Cataloguing-in-Publication Data

Berg, Jeremy Mark
Biochemistry / Jeremy Berg, John Tymoczko, Lubert Stryer.—5th ed.
p. cm.
Fourth ed. by Lubert Stryer.
Includes bibliographical references and index.
ISBN 0-7167-4954-8 (CH1–31 only)—ISBN 0-7167-3051-0 (CH1–34)
ISBN 0-7167-4955-6 (CH32–34 only)—ISBN 0-7167-4684-0 (CH1–34, International edition)
1. Biochemistry. I. Tymoczko, John L. II. Stryer, Lubert. III. Title.

QP514.2 .S66 2001
572—dc21

2001051259

© 2002 by W. H. Freeman and Company; © 1975, 1981, 1988, 1995 by Lubert Stryer. All rights reserved

No part of this book may be reproduced by any mechanical, photographic, or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted, or otherwise copied for public or private use, without written permission from the publisher.

Printed in the United States of America

First printing 2001

W. H. Freeman and Company
41 Madison Avenue, New York, New York 10010
Houndmills, Basingstoke RG21 6XS, England

acids. For example, amino acids such as homoserine and homocysteine tend to form five-membered cyclic forms that limit their use in proteins; the alternative amino acids that are found in proteins—serine and cysteine—do not readily cyclize, because the rings in their cyclic forms are too small (Figure 3.17).

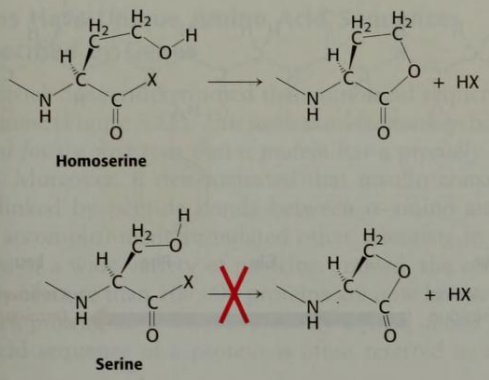


FIGURE 3.17 Undesirable reactivity in amino acids. Some amino acids are unsuitable for proteins because of undesirable cyclization. Homoserine can cyclize to form a stable, five-membered ring, potentially resulting in peptide-bond cleavage. Cyclization of serine would form a strained, four-membered ring and thus is unfavored. X can be an amino group from a neighboring amino acid or another potential leaving group.

3.2 PRIMARY STRUCTURE: AMINO ACIDS ARE LINKED BY PEPTIDE BONDS TO FORM POLYPEPTIDE CHAINS

Proteins are *linear polymers* formed by linking the α-carboxyl group of one amino acid to the α-amino group of another amino acid with a *peptide bond* (also called an *amide bond*). The formation of a dipeptide from two amino acids is accompanied by the loss of a water molecule (Figure 3.18). The equilibrium of this reaction lies on the side of hydrolysis rather than synthesis. Hence, the biosynthesis of peptide bonds requires an input of free energy. Nonetheless, peptide bonds are quite *stable kinetically;* the lifetime of a peptide bond in aqueous solution in the absence of a catalyst approaches 1000 years.

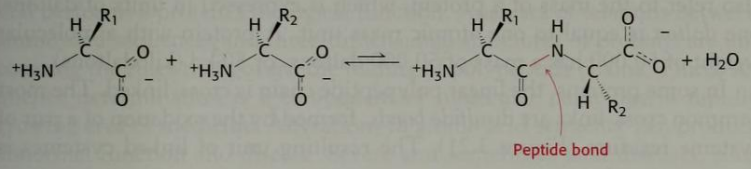


FIGURE 3.18 Peptide-bond formation. The linking of two amino acids is accompanied by the loss of a molecule of water.

A series of amino acids joined by peptide bonds form a *polypeptide chain,* and each amino acid unit in a polypeptide is called a *residue. A polypeptide chain has polarity* because its ends are different, with an α-amino group at one end and an α-carboxyl group at the other. By convention, *the amino end is taken to be the beginning of a polypeptide chain,* and so the sequence of amino acids in a polypeptide chain is written starting with the amino-terminal residue. Thus, in the pentapeptide Tyr-Gly-Gly-Phe-Leu (YGGFL), tyrosine is the amino-terminal (N-terminal) residue and leucine is the carboxyl-terminal (C-terminal) residue (Figure 3.19). Leu-Phe-Gly-Gly-Tyr (LFGGY) is a different pentapeptide, with different chemical properties.

A polypeptide chain consists of a regularly repeating part, called the *main chain* or *backbone,* and a variable part, comprising the distinctive *side chains* (Figure 3.20). The polypeptide backbone is rich in hydrogen-bonding potential. Each residue contains a carbonyl group, which is a good hydrogen-bond acceptor and, with the exception of proline, an NH group, which is a

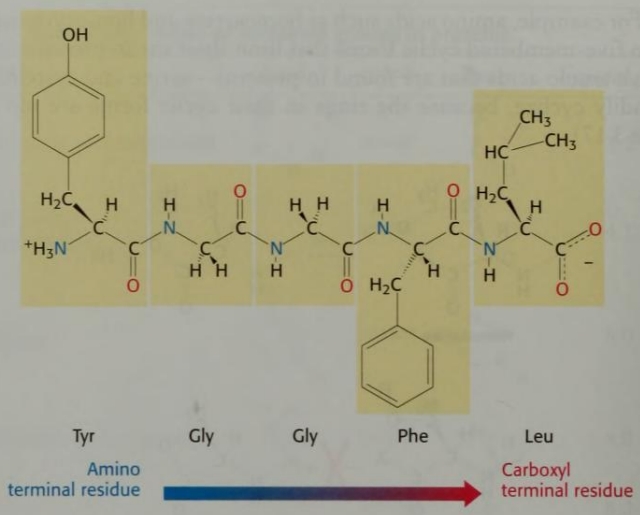


FIGURE 3.19 Amino acid sequences have direction. This illustration of the pentapeptide Try-Gly-Gly-Phe-Leu (YGGFL) shows the sequence from the amino terminus to the carboxyl terminus. This pentapeptide, Leu-enkephalin, is an opioid peptide that modulates the perception of pain. The reverse pentapeptide, Leu-Phe-Gly-Gly-Tyr (LFGGY), is a different molecule and shows no such effects.

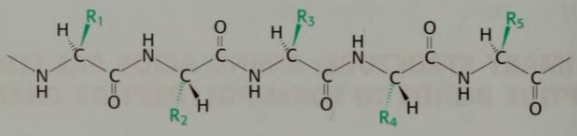


FIGURE 3.20 Components of a polypeptide chain. A polypeptide chain consists of a constant backbone (shown in black) and variable side chains (shown in green).

good hydrogen-bond donor. These groups interact with each other and with functional groups from side chains to stabilize particular structures, as will be discussed in detail.

Most natural polypeptide chains contain between 50 and 2000 amino acid residues and are commonly referred to as *proteins*. Peptides made of small numbers of amino acids are called *oligopeptides* or simply *peptides*. The mean molecular weight of an amino acid residue is about 110, and so the molecular weights of most proteins are between 5500 and 220,000. We can also refer to the mass of a protein, which is expressed in units of daltons; one *dalton* is equal to one atomic mass unit. A protein with a molecular weight of 50,000 has a mass of 50,000 daltons, or 50 kd (kilodaltons).

Dalton—
A unit of mass very nearly equal to that of a hydrogen atom. Named after John Dalton (1766–1844), who developed the atomic theory of matter.

Kilodalton (kd)—
A unit of mass equal to 1000 daltons.

In some proteins, the linear polypeptide chain is cross-linked. The most common cross-links are *disulfide bonds*, formed by the oxidation of a pair of cysteine residues (Figure 3.21). The resulting unit of linked cysteines is

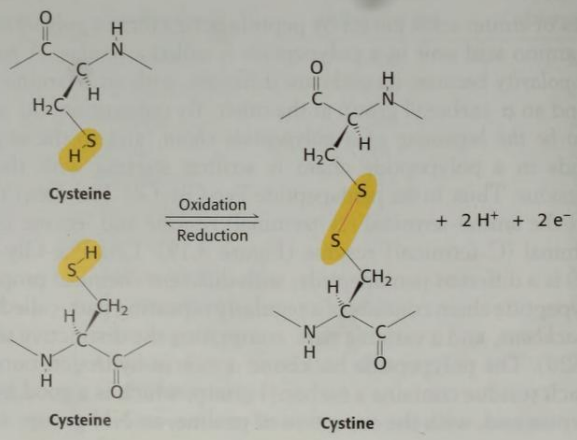


FIGURE 3.21 Cross-links. The formation of a disulfide bond from two cysteine residues is an oxidation reaction.

called *cystine*. Extracellular proteins often have several disulfide bonds, whereas intracellular proteins usually lack them. Rarely, nondisulfide cross-links derived from other side chains are present in some proteins. For example, collagen fibers in connective tissue are strengthened in this way, as are fibrin blood clots.

3.2.1 Proteins Have Unique Amino Acid Sequences That Are Specified by Genes

In 1953, Frederick Sanger determined the amino acid sequence of insulin, a protein hormone (Figure 3.22). *This work is a landmark in biochemistry because it showed for the first time that a protein has a precisely defined amino acid sequence*. Moreover, it demonstrated that insulin consists only of L amino acids linked by peptide bonds between α-amino and α-carboxyl groups. This accomplishment stimulated other scientists to carry out sequence studies of a wide variety of proteins. Indeed, the complete amino acid sequences of more than 100,000 proteins are now known. *The striking fact is that each protein has a unique, precisely defined amino acid sequence*. The amino acid sequence of a protein is often referred to as its *primary structure*.

A series of incisive studies in the late 1950s and early 1960s revealed that the amino acid sequences of proteins are genetically determined. The sequence of nucleotides in DNA, the molecule of heredity, specifies a complementary sequence of nucleotides in RNA, which in turn specifies the amino acid sequence of a protein. In particular, each of the 20 amino acids of the repertoire is encoded by one or more specific sequences of three nucleotides (Section 5.5).

Knowing amino acid sequences is important for several reasons. First, knowledge of the sequence of a protein is usually essential to elucidating its mechanism of action (e.g., the catalytic mechanism of an enzyme). Moreover, proteins with novel properties can be generated by varying the sequence of known proteins. Second, amino acid sequences determine the three-dimensional structures of proteins. Amino acid sequence is the link between the genetic message in DNA and the three-dimensional structure that performs a protein's biological function. Analyses of relations between amino acid sequences and three-dimensional structures of proteins are uncovering the rules that govern the folding of polypeptide chains. Third, sequence determination is a component of molecular pathology, a rapidly growing area of medicine. Alterations in amino acid sequence can produce abnormal function and disease. Severe and sometimes fatal diseases, such as sickle-cell anemia and cystic fibrosis, can result from a change in a single amino acid within a protein. Fourth, the sequence of a protein reveals much about its evolutionary history (see Chapter 7). Proteins resemble one another in amino acid sequence only if they have a common ancestor. Consequently, molecular events in evolution can be traced from amino acid sequences; molecular paleontology is a flourishing area of research.

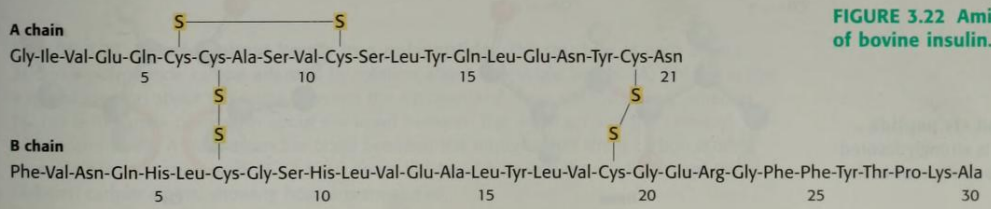


FIGURE 3.22 Amino acid sequence of bovine insulin.