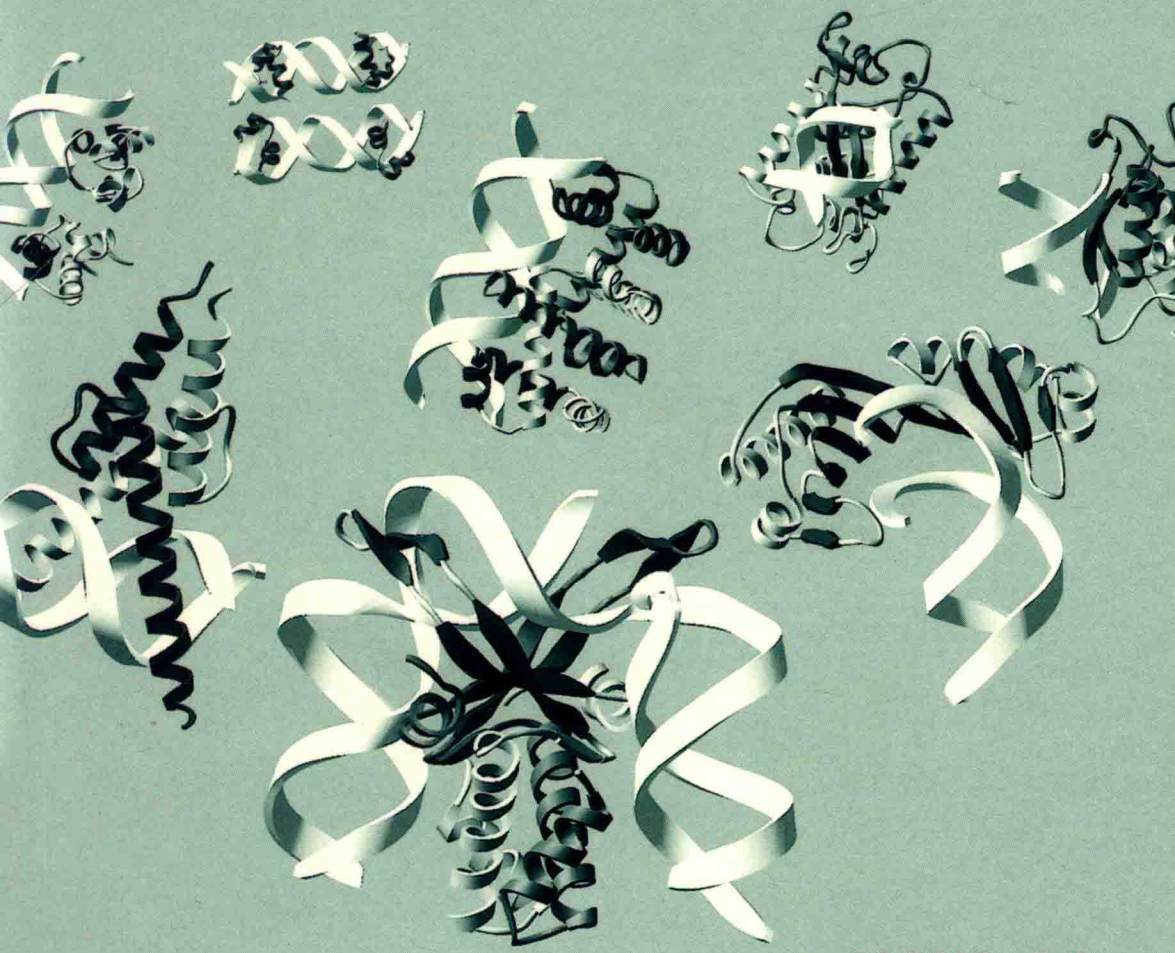
BIOORGANIC CHEMISTRY

PEPTIDES AND PROTEINS



EDITED BY

SIDNEY M. HECHT

Bioorganic Chemistry: Peptides and Proteins

Edited by

Sidney M. Hecht
University of Virginia

New York *Oxford*
OXFORD UNIVERSITY PRESS
1998

Oxford University Press

Oxford New York
Athens Auckland Bangkok Bogota Bombay Buenos Aires
Calcutta Cape Town Dar es Salaam Delhi Florence Hong Kong
Istanbul Karachi Kuala Lumpur Madras Madrid Melbourne
Mexico City Nairobi Paris Singapore Taipei Tokyo Toronto Warsaw

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.,
198 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Bioorganic chemistry : peptides and proteins / edited by Sidney M.
Hecht.
p. cm. — (Topics in bioorganic and biological chemistry)
Includes bibliographical references and index.
ISBN 0-19-508468-3 (cloth : alk. paper)
1. Proteins. 2. Peptides. I. Hecht, Sidney M. II. Series.
QP551.B474 1998
572.6—dc21 97-27060

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Bioorganic Chemistry: Peptides and Proteins

Topics in Bioorganic and Biological Chemistry
A Series of Books in Support of Teaching and Research

Series Editors:
Sidney M. Hecht, University of Virginia
Richard L. Schowen, University of Kansas

Bioorganic Chemistry: Nucleic Acids
S. Hecht, editor

Bioorganic Chemistry: Peptides and Proteins
S. Hecht, editor

Bioorganic Chemistry: Carbohydrates
S. Hecht, editor

Enzyme Catalysis
R. Schowen

Steady State Enzyme Kinetics
P. Cook

Preface

This is the second volume of a series of books in bioorganic and biological chemistry. The first volume, dealing with nucleic acids, appeared in the spring of 1996. A third volume in the area of carbohydrates is in production and will appear shortly.

As noted in the preface to the first volume, the increasingly detailed understanding of the molecular basis by which biological systems operate has dramatically increased the range of studies now considered to be within the domain of organic chemistry. The "core" of expertise required to function effectively in this rapidly expanding field has increased correspondingly. This poses an ongoing educational challenge both for scientists presently working in the field and, especially, for students encountering the subject matter for the first time. This series of books is intended to support the teaching of graduate students in bioorganic chemistry.

In keeping with the format of the first volume, *Bioorganic Chemistry: Peptides and Proteins* consists of a set of 14 chapters, approximately equal to the number of weeks in a semester. The subject matter of each chapter is judged to be both representative of and central to an understanding of ongoing research activity in the field of peptides and proteins, as practiced by bioorganic chemists. Each chapter begins with an overview of basic principles and a summary of key findings that form the basis for current research activity in the specific area of focus in the chapter. The remainder of each chapter presents a limited number of examples of recent studies in greater depth. The chapters are thus organized in much the same fashion as typical lectures in special topics courses. A set of overheads corresponding to each of the figures in the book is available to aid classroom presentation.

In addition to my own favorable experience in teaching the subject matter of all three books in the set to graduate students at the University of Virginia, others who have used the nucleic acids book for a course have affirmed that it functions as intended. Gratifyingly, numerous colleges have commented that the book has been no less valuable as an educational tool within their own research laboratories.

I would like to thank the authors of this volume for their efforts in writing the chapters. Oxford University Press has continued to provide excellent support for this educational experiment; the ongoing help and advice of Bob Rogers, Senior Editor, has been invaluable. This volume was typed by Vickie Thomas, who provided uncommon technical support and much patience. Carolyn Esau assisted with verification of numerous literature references. I thank them both for their assistance. I also acknowledge with gratitude the contributions of many graduate and postdoctoral students who have participated enthusiastically in the special topic courses that I have given at the University of Virginia based on the material in this book.

Charlottesville
October, 1997

S. M. H.

Contributors

Milton J. Axley
Naval Medical Research Institute

Paul R. Blake
University of Maryland, Baltimore County

Charles W. Carter, Jr.
University of North Carolina, Chapel Hill

Michael C. Fitzgerald
The Scripps Research Institute

Maurille J. Fournier
University of Massachusetts, Amherst

John A. Gerlt
University of Maryland

Lila M. Gierasch
University of Massachusetts, Amherst

Ronald A. Hernan
The Scripps Research Institute

Donald Hilvert
The Scripps Research Institute

Victor J. Hruby
University of Arizona

Michael Kahn
Molecumetics

Stephen B. H. Kent
The Scripps Research Institute

Zhi-Ping Liu
University of Massachusetts, Amherst

Paul J. Loida
University of Illinois at Urbana-Champaign

Gavin MacBeath
The Scripps Research Institute

Thomas L. Mason
University of Massachusetts, Amherst

Michael A. Massiah
University of Maryland, Baltimore County

Jean-Philippe Meyer
University of Arizona

Hiroshi Nakanishi
Molecumetics

Josep Rizo
University of Texas, Southwestern

Alan J. Russell
University of Pittsburgh

Jumi A. Shin
The Scripps Research Institute

John E. Shively
Beckman Research Institute

Stephen G. Sligar
University of Illinois at Urbana-Champaign

Arno F. Spatola
University of Louisville

Michael F. Summers
University of Maryland, Baltimore County

David A. Tirrell
University of Massachusetts, Amherst

Jane G. Tirrell
University of Massachusetts, Amherst

Zhen Yang
University of Pittsburgh

Bioorganic Chemistry: Peptides and Proteins

Contents

1

Introduction to Peptides and Proteins

Milton J. Axley

Proteins and peptides provide many of the chemical and physical processes that constitute life in biological organisms. These molecules are ubiquitous in all living creatures. Although certain types of proteins are shared by all organisms, each life form has a unique make-up of proteins that defines the characteristics of that organism. The synthesis of a protein within an organism is determined by the genetic make-up of the organism, and most of the genetic make-up of an organism is devoted to the expression of proteins.

The four main groupings of biological macromolecules each have their own functionalities. *Nucleic acids* store and manipulate the genetic information that details the make-up of an organism. *Carbohydrates* store energy and, in plants, serve structural purposes. *Lipid* assemblies make up the bulk of cellular membranes that compartmentalize cells and subcellular organelles. *Proteins* and *peptides* are the agents of action in a cell. Some of the biological roles in which proteins function include: chemical reaction catalysis, motility, physiological regulation, transport, structural composition, and defense. Proteins perform most of the activities and provide much of the structure that constitutes cells and organisms.

The Dutch chemist Gerardus Mulder coined the name "protein" in 1838, when he found proteins to be the major chemical constituent of cellular matter. The name is derived from the Greek word "proteis," which means "of primary importance." Proteins have more than lived up to their name throughout all the advances in understanding since Mulder's time, and today these molecules continue to be the center point for progress in diverse areas of scientific and industrial research.

Modern biotechnology is directed toward the exploitation of proteins and peptides. The ability to harness the vast potential of proteins provides a powerful tool for medicine, as proteins are involved in and can be used for the treatment of infectious, genetic, and traumatic diseases. Agriculture benefits from modern protein chemistry with improved crop yields, pest control, and animal husbandry. Chemistry has gained new and more efficient synthetic methods catalyzed by proteins.

Structure

Proteins and peptides are linear polymers consisting of amino acids. Short chains of amino acids (two to about forty) are known as peptides. Two amino acids linked by a peptide bond constitutes a dipeptide, while three amino acids linked together make up

a tripeptide. Longer chains containing more than 30 to 50 amino acids are polypeptides. There is no single cutoff length that separates peptides from polypeptides; a 40-amino acid chain might be called either a peptide or a polypeptide. The individual amino acids within a peptide or polypeptide are called subunits or residues.

Proteins are polypeptides with a function. The terms "protein" and "polypeptide" are sometimes used interchangeably, although polypeptide strictly refers to an amino acid chain and protein refers to the overall entity, possibly including multiple polypeptide subunits and cofactors in a functional conformation. Enzymes are proteins that catalyze chemical reactions.

Amino Acids

In order to understand the nature of polypeptides, it is necessary to understand the characteristics of the amino acid subunits. Amino acids contain an amino group, a carboxy group, a hydrogen atom, and a side group all connected to a central (α) carbon atom (see Fig. 1-1). Since the central carbon is attached to four different groups, amino acids (except for glycine) are chiral molecules. All amino acids found in proteins and nearly all found in peptides are of the "L" (or S in the Cahn-Ingold-Prelog nomenclature) configuration. Amino acids in solution generally exist as zwitterions, as their amino and carboxy groups are both charged at pH values between 3 and 9.

As amino and carboxy groups are common to all amino acids, it is the side group that defines amino acids and determines their individual characteristics. The chemical and structural natures of the side groups determine the structure and function of the amino acid within the context of a protein or peptide. A variety of functional groups are available as side chains, including acidic, basic, hydrophobic, hydrophilic, redox-active, bulky, and compact chemical moieties. This range of functional groups provides the chemical basis for the vast array of biochemical properties displayed by proteins and peptides.

Twenty amino acids are commonly found as constituents of proteins and peptides in all living organisms. Figure 1-2 gives diagrams of the twenty "standard amino acids." Amino acids can be grouped into several types, based on the chemical properties of their side groups. Amino acids are often described as acidic, basic, polar, nonpolar, bulky, and so on. Other groupings are obviously also possible. Such descriptions based on side group properties are informative within the context of a polypeptide chain of amino acids, but may be inaccurate for the individual free amino acid. For example, alanine may be called a nonpolar or hydrophobic amino acid due to its methyl side group, but the free amino acid alanine by itself will have two charged groups at neutral pH due to the carboxy and amino groups. Groupings of amino acids allow comparison of the side group properties and prediction of the effect that amino acid sub-

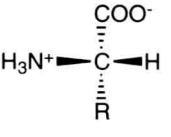


Figure 1-1. Structure of an L-amino acid indicating absolute stereochemistry. The general structure of an L-amino acid consists of a central carbon atom surrounded by four different chemical groups: carboxylic acid and amino groups, a hydrogen atom, and a side chain. The side chain is here represented by R.

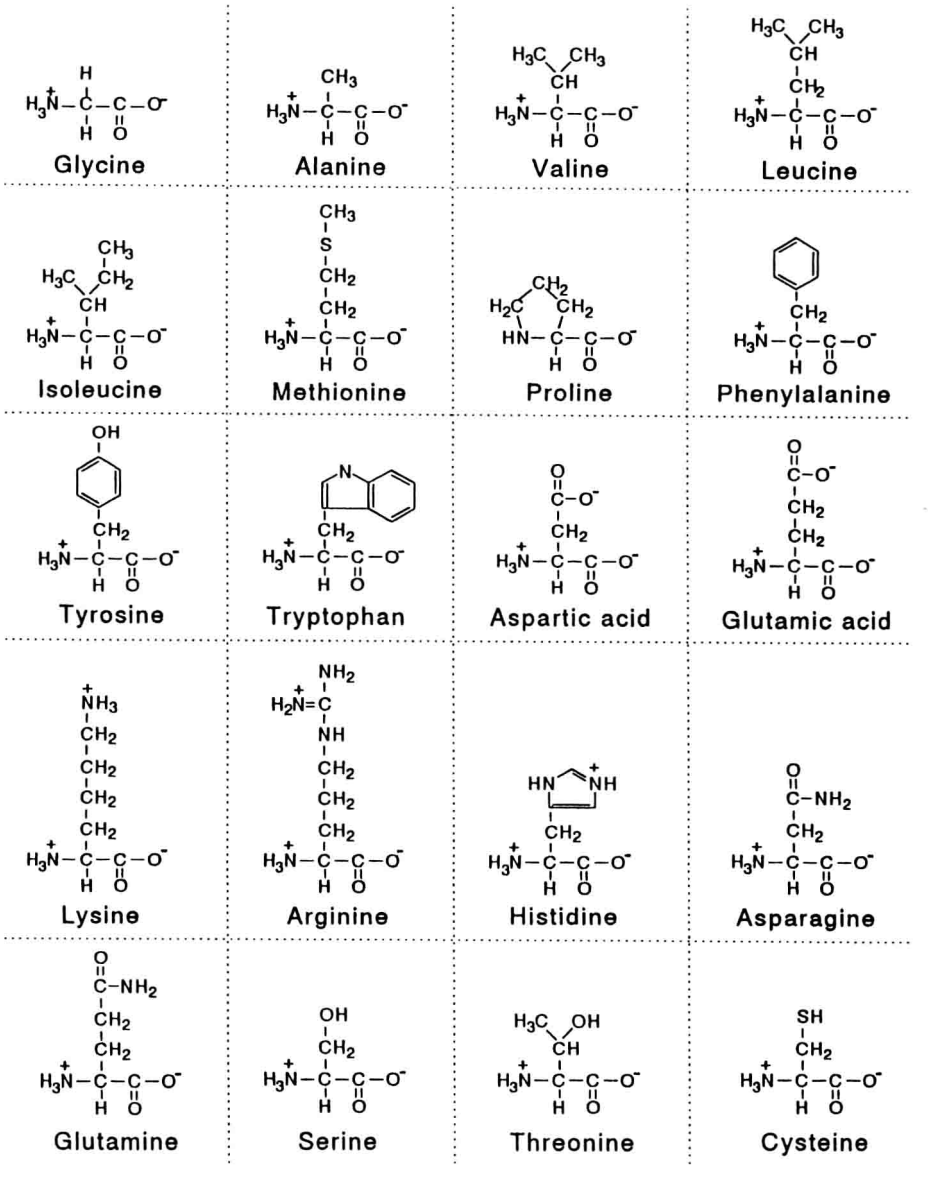


Figure 1-2. Chemical structures for the twenty common amino acids. The amino acids are presented in the order in which they are discussed in the text. Ionic charges reflect the situations at pH 7.

stitutions may have on the structure or biochemistry of polypeptides. For example, substitution of the amino acid aspartic acid for glutamic acid at the same position in a polypeptide might be expected to have little effect on the physical-chemical properties, as both amino acids contain carboxylic acid side groups. However, substitution of tryptophan, containing a hydrophobic side group, at the same position might have a more drastic effect on polypeptide structure and function.

Glycine is the smallest and simplest amino acid, as it has only a single hydrogen atom as a side group. Since the central carbon of glycine has two hydrogens attached to it, this is the only standard amino acid that is not chiral.

There are six amino acids with small, nonpolar side groups. Alanine has a methyl side chain, valine an isopropyl, leucine a 2-methylpropyl and isoleucine a 1-methylpropyl group. Methionine is one of two sulfur-containing amino acids, with a methylated thioether side group. Proline is unusual among the 20 standard amino acids, because its side chain is attached to the amino group to form a five-membered ring. The nitrogen group of proline is actually a secondary amine.

Three amino acids have bulky, uncharged, aromatic side groups. Phenylalanine has a benzyl functional group, while tyrosine has a phenol, and tryptophan has an indole moiety in its side group. Phenylalanine and tryptophan are considered nonpolar and hydrophobic, but the hydroxyl group of tyrosine gives its side chain some polar character.

Five amino acids have R groups that are charged at neutral pH. Aspartic acid and glutamic acid both have carboxyl-containing side groups, and they are, therefore, organic acids. Lysine, arginine, and histidine have basic side groups containing amino, guanidino, and imidazolium functional groups, respectively. The side group of histidine has a pK_a of about 6, and is the only amino acid of this group with a pK_a near physiological pH.

The side groups of five amino acids are small and polar. Asparagine and glutamine are the amide forms of aspartic and glutamic acids, respectively. Serine has a hydroxymethyl R group, and threonine has a 1-hydroxyethyl side group. Cysteine is a sulfur-containing amino acid, as its side group contains a thiol (-SH) functional group. The thiol of cysteine readily undergoes a reversible oxidation, and this reaction is important to many structural and functional properties of polypeptides. Specifically, the thiols of two cysteines can be oxidized to form a covalent disulfide bond between the two amino acids. The resulting derivatized amino acid, called cystine, stabilizes the three-dimensional structure of many polypeptides.

Amino acids are often referred to by three- or one-letter abbreviations. These are particularly useful for writing out the sequences of polypeptides. Table 1-1 presents the accepted three-letter and one-letter abbreviations for the 20 standard amino acids.

In addition to the 20 standard amino acids, several other amino acids are found naturally in certain proteins and peptides. In most such cases, amino acids are chemically modified after incorporation into a polypeptide. This posttranslational modification can be essential to the function of the polypeptide. For example, tropocollagen, the polypeptide component of collagen, contains a relatively large proportion (about 9%) of the amino acid hydroxyproline. The hydroxyproline is formed by hydroxylation of proline after synthesis of the polypeptide chain by an enzyme that requires ascorbic acid (vitamin C) for its activity. Without the activity of the hydroxylating enzyme, collagen is not properly formed. The human disease scurvy, characterized by improper collagen formation, is caused by vitamin C deficiency.

A number of proteins have been shown to contain phosphorylated derivatives of the hydroxyl-containing amino acids tyrosine, serine, and threonine. In these derivatives, a phosphate group is attached to the hydroxyl group as an ester to form phosphotyrosine, phosphoserine, and phosphothreonine, respectively. These phosphoester derivatives are formed posttranslationally, that is, after synthesis of the polypeptide by translation within a cell. Phosphorylation occurs at specific sites of certain proteins, and the reaction is reversible. Specific phosphorylation and dephosphorylation of certain proteins affect the activities of those proteins: increasing, decreasing, or otherwise alter-

Table 1-1 Amino Acid Abbreviations

Amino Acid	Three-letter Code	One-letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

ing the activities of the proteins. Thus, phosphorylation of proteins can act as a reversible control mechanism, switching proteins between active and inactive forms.

Most of the nonstandard amino acids found in natural proteins are produced by post-translational modification of one of the twenty standard amino acids present in a peptide chain. However, this is not the case for selenocysteine, which is identical to cysteine except that a selenium atom replaces the sulfur atom of cysteine. Selenocysteine is found as an essential component of certain proteins in many classes of organisms. It is coded for in the DNA, and a specific tRNA directs its cotranslational insertion into proteins. For these reasons, selenocysteine has been called the "twenty-first amino acid" (Söll, 1988). In mammals, selenium is a required nutrient, and at least half a dozen proteins contain essential selenocysteine residues. Selenium deficiency causes a cardiomyopathy in humans and white-muscle disease in domestic livestock. Glutathione peroxidase is a mammalian selenoprotein that catalyzes the repair of cellular oxidative damage caused by free radicals and peroxides; therefore, this selenoprotein may play a role in prevention of cancer and aging.

Polypeptide Synthesis

Each polypeptide synthesized in a cell is encoded by a single gene. The DNA sequences of chromosomal genes directly encode the amino acid sequences of all proteins and most peptides. The information encoding each amino acid in a polypeptide is found in

a "codon." The codons of a gene are arranged sequentially in the same order in which the amino acids appear in a polypeptide.

The linear arrangement of information found in the nucleotide sequence of DNA can be directly translated into the linear arrangement of amino acids in a polypeptide sequence through the use of the universal *genetic code*. The genetic code is thus a biological algorithm for the translation of information found in gene codons into amino acids in polypeptides.

A codon consists of three adjacent nucleotides within a gene, and the three nucleotides of a codon are sometimes called a triplet. Since there are four different nucleotide bases in DNA (and RNA), there are 64 (4^3) possible triplet codons. As there are only 20 standard amino acids for which to code, there is redundancy in the genetic code. That is, there is more than one codon for some of the amino acids. In addition, three of the codons do not code for amino acids. These codons are stop codons, also known as end or termination codons. Stop codons designate the end of the polypeptide coding sequence of a gene.

The genetic code is said to be universal. Virtually every tested species of living organism uses the universal genetic code. However, there are also exceptions where gene codons do not match the corresponding amino acids found in polypeptides. One type of exception is when the amino acid is chemically altered after polypeptide synthesis. Examples of this "posttranslational modification" include glycosylated and phosphorylated amino acids.

In other exceptions to the universal genetic code, the codons that normally code for termination instead code for amino acids. For example, in *Tetrahymena* the UAA and UAG codons designate glutamine insertion into polypeptide; in *Mycoplasma* and eukaryotic mitochondria the UGA codon specifies tryptophan, while in many eukaryotes and prokaryotes the UGA codon can also code for selenocysteine.

Synthesis of a polypeptide by a living cell occurs in two discrete steps, transcription and translation. In the first step, transcription, an RNA (ribonucleic acid) copy of the coding portion of the DNA (deoxyribonucleic acid) gene is synthesized. RNA consists of a linear array of four different ribonucleotide bases: adenine, uracil, guanine, and cytosine, abbreviated A, U, G, and C, respectively. DNA consists of a string of deoxyribonucleotides, which also contain the bases adenine, guanine, and cytosine; however, DNA contains thymine in place of uracil.

The RNA copy that is transcribed from the DNA gene is called the messenger RNA, or mRNA. Although DNA is the storehouse for coding information, it is actually the mRNA that is used during translation, the second major step of protein biosynthesis. During translation, the codons contained in the mRNA are translated sequentially via the genetic code, affording a specific polypeptide sequence. Two other macromolecular species are required for translation, namely transfer RNA and ribosomes. Transfer RNA (tRNA) molecules carry amino acids to the site of translation, which is within the ribosome. Each tRNA has a single anticodon, a three-base sequence complementary to an mRNA codon. There is a tRNA for every codon, and tRNAs are linked to (activated with) the amino acid corresponding with the tRNA anticodon.

The tRNA anticodon can anneal to (hybridize with) the corresponding codon through nucleotide base pairing interactions within the context of the ribosome. Ribosomes are very large macromolecular bodies, made up of numerous proteins and a few large RNA molecules (ribosomal RNA). During translation, ribosomes bind the mRNA, direct the codon-anticodon interactions between the mRNA and tRNAs and then link together the amino acids presented by the tRNAs.