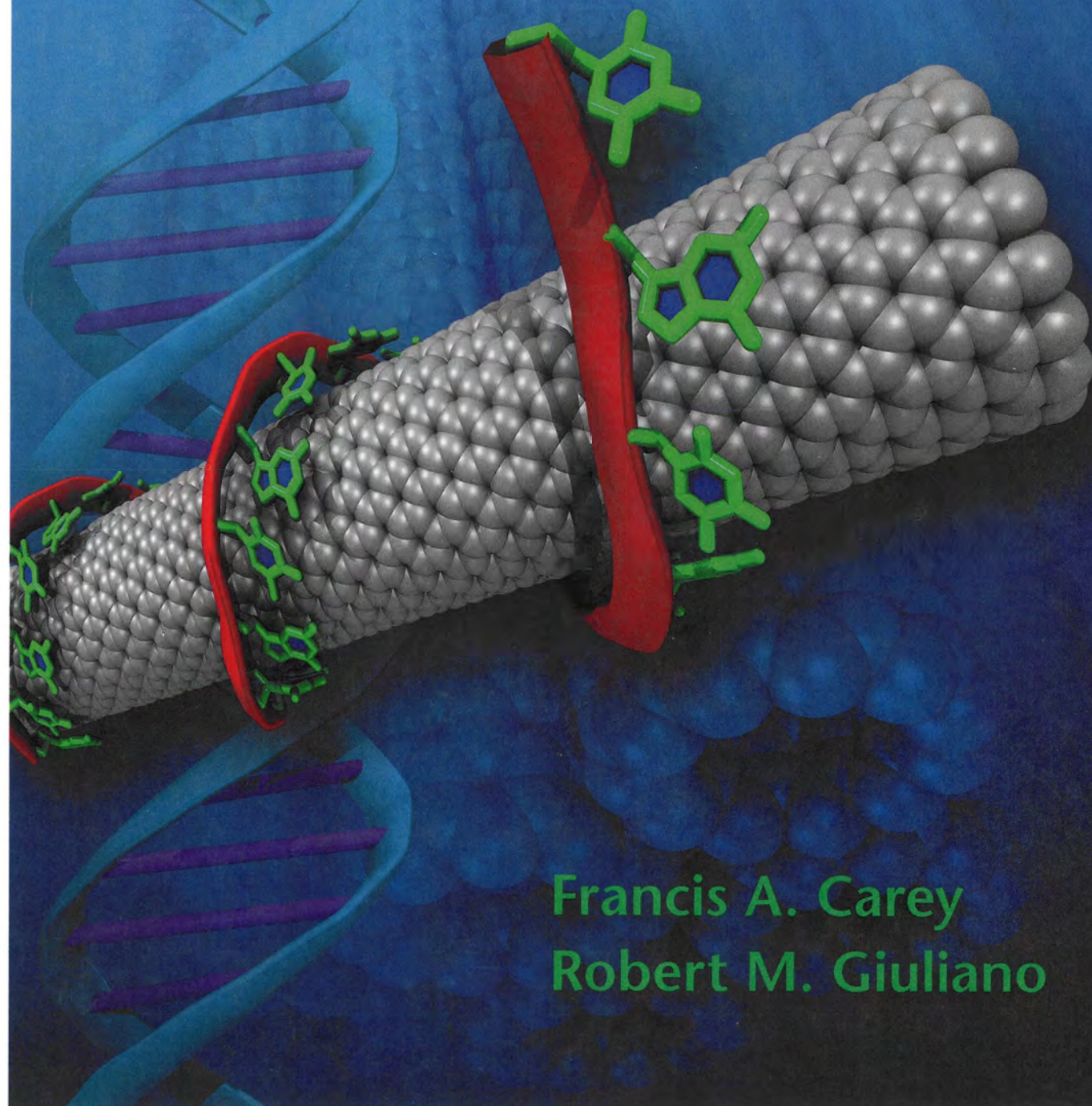


eighth edition

# Organic Chemistry



Francis A. Carey  
Robert M. Giuliano

Eighth Edition

# Organic Chemistry

**Francis A. Carey**  
University of Virginia

**Robert M. Giuliano**  
Villanova University





ORGANIC CHEMISTRY, EIGHTH EDITION

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2011 by The McGraw-Hill Companies, Inc. All rights reserved. Previous editions © 2008, 2006, and 2003. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOW/DOW 1 0 9 8 7 6 5 4 3 2 1 0

ISBN 978-0-07-340261-1

MHID 0-07-340261-3

Publisher: *Ryan Blankenship*

Senior Sponsoring Editor: *Tamara L. Hodge*

Vice-President New Product Launches: *Michael Lange*

Developmental Editor: *Jodi Rhomberg*

Senior Marketing Manager: *Todd L. Turner*

Lead Project Manager: *Sheila M. Frank*

Lead Production Supervisor: *Sandy Ludovissy*

Lead Media Project Manager: *Judi David*

Senior Designer: *Laurie B. Janssen*

Cover/Interior Designer: *Christopher Reese*

Cover Illustration: *Robert R. Johnson, Johnson Group, University of Pennsylvania*

Senior Photo Research Coordinator: *John C. Leland*

Photo Research: *David Tietz/Editorial Image, LLC*

Supplement Producer: *Mary Jane Lampe*

Compositor: *Aptara®, Inc.*

Typeface: 10.5/12 Times LT Std

Printer: *R. R. Donnelley*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

Library of Congress Cataloging-in-Publication Data

Carey, Francis A., 1937-

Organic chemistry / Francis A. Carey.—8th ed. / Robert M. Giuliano.

p. cm.

Includes index.

ISBN 978-0-07-340261-1—ISBN 0-07-340261-3 (hard copy : alk. paper)

I. Chemistry, Organic—Textbooks. I. Giuliano, Robert M., 1954- II. Title.

QD251.3.C37 2011

547—dc22

2009034746

To our wives, Margot  
Giuliano and Jill Carey,  
and our children  
Michael, Ellen, and  
Christopher Giuliano  
and Andrew, Robert,  
and William Carey.

*Frank Carey*

*Bob Giuliano*

# Brief Contents

|                                                                              |      |
|------------------------------------------------------------------------------|------|
| List of Important Features                                                   | xvii |
| Preface                                                                      | xxii |
| Acknowledgments                                                              | xxx  |
| 1 Structure Determines Properties                                            | 2    |
| 2 Alkanes and Cycloalkanes: Introduction to Hydrocarbons                     | 56   |
| 3 Alkanes and Cycloalkanes: Conformations and cis-trans Stereoisomers        | 100  |
| 4 Alcohols and Alkyl Halides                                                 | 137  |
| 5 Structure and Preparation of Alkenes: Elimination Reactions                | 184  |
| 6 Addition Reactions of Alkenes                                              | 226  |
| 7 Stereochemistry                                                            | 278  |
| 8 Nucleophilic Substitution                                                  | 322  |
| 9 Alkynes                                                                    | 359  |
| 10 Conjugation in Alkadienes and Allylic Systems                             | 388  |
| 11 Arenes and Aromaticity                                                    | 428  |
| 12 Reactions of Arenes: Electrophilic and Nucleophilic Aromatic Substitution | 478  |
| 13 Spectroscopy                                                              | 538  |
| 14 Organometallic Compounds                                                  | 606  |
| 15 Alcohols, Diols, and Thiols                                               | 646  |
| 16 Ethers, Epoxides, and Sulfides                                            | 686  |
| 17 Aldehydes and Ketones: Nucleophilic Addition to the Carbonyl Group        | 724  |
| 18 Carboxylic Acids                                                          | 776  |
| 19 Carboxylic Acid Derivatives: Nucleophilic Acyl Substitution               | 812  |
| 20 Enols and Enolates                                                        | 866  |
| 21 Amines                                                                    | 930  |
| 22 Phenols                                                                   | 988  |
| 23 Carbohydrates                                                             | 1022 |
| 24 Lipids                                                                    | 1074 |
| 25 Amino Acids, Peptides, and Proteins                                       | 1116 |
| 26 Nucleosides, Nucleotides, and Nucleic Acids                               | 1174 |
| 27 Synthetic Polymers                                                        | 1216 |
| Glossary                                                                     | G-1  |
| Credits                                                                      | C-1  |
| Index                                                                        | I-1  |

# Contents

List of Important Features xvii

Preface xxii

Acknowledgments xxx

## CHAPTER 1

### Structure Determines Properties 2

- 1.1 Atoms, Electrons, and Orbitals 3
- 1.2 Ionic Bonds 6
- 1.3 Covalent Bonds, Lewis Structures, and the Octet Rule 8
- 1.4 Double Bonds and Triple Bonds 10
- 1.5 Polar Covalent Bonds, Electronegativity, and Bond Dipoles 11
  - Electrostatic Potential Maps 13
- 1.6 Formal Charge 13
- 1.7 Structural Formulas of Organic Molecules 16
- 1.8 Resonance 19
- 1.9 Writing Organic Structures 23
- 1.10 The Shapes of Some Simple Molecules 26
  - Molecular Modeling 26
- 1.11 Molecular Dipole Moments 28
- 1.12 Curved Arrows and Chemical Reactions 29
- 1.13 Acids and Bases: The Arrhenius View 32
- 1.14 Acids and Bases: The Brønsted-Lowry View 33
- 1.15 What Happened to  $pK_b$ ? 37
- 1.16 How Structure Affects Acid Strength 38
- 1.17 Acid-Base Equilibria 42
- 1.18 Lewis Acids and Lewis Bases 45
- 1.19 Summary 46
  - Problems 49
  - Descriptive Passage and Interpretive Problems 1: Amide Lewis Structures 55

## CHAPTER 2

### Alkanes and Cycloalkanes: Introduction to Hydrocarbons 56

- 2.1 Classes of Hydrocarbons 57
- 2.2 Electron Waves and Chemical Bonds 58
- 2.3 Bonding in  $H_2$ : The Valence Bond Model 59
- 2.4 Bonding in  $H_2$ : The Molecular Orbital Model 60

- 2.5 Introduction to Alkanes: Methane, Ethane, and Propane 62
  - Methane and the Biosphere 63
- 2.6  $sp^3$  Hybridization and Bonding in Methane 63
- 2.7 Bonding in Ethane 65
- 2.8 Isomeric Alkanes: The Butanes 65
- 2.9 Higher  $n$ -Alkanes 66
- 2.10 The  $C_5H_{12}$  Isomers 67
- 2.11 IUPAC Nomenclature of Unbranched Alkanes 69
  - What's in a Name? Organic Nomenclature 70
- 2.12 Applying the IUPAC Rules: The Names of the  $C_6H_{14}$  Isomers 71
- 2.13 Alkyl Groups 72
- 2.14 IUPAC Names of Highly Branched Alkanes 74
- 2.15 Cycloalkane Nomenclature 75
- 2.16 Sources of Alkanes and Cycloalkanes 76
- 2.17 Physical Properties of Alkanes and Cycloalkanes 78
- 2.18 Chemical Properties: Combustion of Alkanes 80
  - Thermochemistry 82
- 2.19 Oxidation-Reduction in Organic Chemistry 83
- 2.20  $sp^2$  Hybridization and Bonding in Ethylene 85
- 2.21  $sp$  Hybridization and Bonding in Acetylene 87
- 2.22 Bonding in Water and Ammonia: Hybridization of Oxygen and Nitrogen 89
- 2.23 Which Theory of Chemical Bonding Is Best? 90
- 2.24 Summary 91
  - Problems 95
  - Descriptive Passage and Interpretive Problems 2: Some Biochemical Reactions of Alkanes 99

## CHAPTER 3

### Alkanes and Cycloalkanes: Conformations and cis-trans Stereoisomers 100

- 3.1 Conformational Analysis of Ethane 102
- 3.2 Conformational Analysis of Butane 105
  - Molecular Mechanics Applied to Alkanes and Cycloalkanes 107
- 3.3 Conformations of Higher Alkanes 108
- 3.4 The Shapes of Cycloalkanes: Planar or Nonplanar? 108
- 3.5 Small Rings: Cyclopropane and Cyclobutane 109
- 3.6 Cyclopentane 110
- 3.7 Conformations of Cyclohexane 111
- 3.8 Axial and Equatorial Bonds in Cyclohexane 112
- 3.9 Conformational Inversion in Cyclohexane 114

- 3.10 Conformational Analysis of Monosubstituted Cyclohexanes 115  
Enthalpy, Free Energy, and Equilibrium Constant 118
- 3.11 Disubstituted Cyclohexanes: cis-trans Stereoisomers 119
- 3.12 Conformational Analysis of Disubstituted Cyclohexanes 120
- 3.13 Medium and Large Rings 124
- 3.14 Polycyclic Ring Systems 124
- 3.15 Heterocyclic Compounds 127
- 3.16 **Summary** 128  
**Problems** 131  
**Descriptive Passage and Interpretive Problems 3:**  
**Cyclic Forms of Carbohydrates** 136

## CHAPTER 4

### Alcohols and Alkyl Halides 137

- 4.1 Functional Groups 138
- 4.2 IUPAC Nomenclature of Alkyl Halides 140
- 4.3 IUPAC Nomenclature of Alcohols 141
- 4.4 Classes of Alcohols and Alkyl Halides 141
- 4.5 Bonding in Alcohols and Alkyl Halides 142
- 4.6 Physical Properties of Alcohols and Alkyl Halides: Intermolecular Forces 143
- 4.7 Preparation of Alkyl Halides from Alcohols and Hydrogen Halides 147
- 4.8 Mechanism of the Reaction of Alcohols with Hydrogen Halides: Hammond's Postulate 148  
Mechanism 4.1 Formation of *tert*-Butyl Chloride from *tert*-Butyl Alcohol and Hydrogen Chloride 149
- 4.9 Potential Energy Diagrams for Multistep Reactions: The  $S_N1$  Mechanism 153
- 4.10 Structure, Bonding, and Stability of Carbocations 154
- 4.11 Effect of Alcohol Structure on Reaction Rate 157
- 4.12 Reaction of Methyl Primary Alcohols with Hydrogen Halides: The  $S_N2$  Mechanism 158  
Mechanism 4.2 Formation of 1-Bromoheptane from 1-Heptanol and Hydrogen Bromide 159
- 4.13 More on Activation Energy 160
- 4.14 Other Methods for Converting Alcohols to Alkyl Halides 160  
Mechanism 4.3 Conversion of an Alcohol to an Alkyl Chloride with Thionyl Chloride 161
- 4.15 Halogenation of Alkanes 162
- 4.16 Chlorination of Methane 162
- 4.17 Structure and Stability of Free Radicals 163  
From Bond Energies to Heats of Reaction 167
- 4.18 Mechanism of Methane Chlorination 168  
Mechanism 4.4 Free-Radical Chlorination of Methane 168
- 4.19 Halogenation of Higher Alkanes 170

- 4.20 **Summary** 174  
**Problems** 177  
**Descriptive Passage and Interpretive Problems 4:**  
**More About Potential Energy Diagrams** 182

## CHAPTER 5

### Structure and Preparation of Alkenes: Elimination Reactions 184

- 5.1 Alkene Nomenclature 185
- 5.2 Structure and Bonding in Alkenes 187  
Ethylene 188
- 5.3 Isomerism in Alkenes 189
- 5.4 Naming Stereoisomeric Alkenes by the *E-Z* Notational System 190
- 5.5 Physical Properties of Alkenes 192
- 5.6 Relative Stabilities of Alkenes 194
- 5.7 Cycloalkenes 197
- 5.8 Preparation of Alkenes: Elimination Reactions 198
- 5.9 Dehydration of Alcohols 199
- 5.10 Regioselectivity in Alcohol Dehydration: The Zaitsev Rule 200
- 5.11 Stereoselectivity in Alcohol Dehydration 202
- 5.12 The  $E1$  and  $E2$  Mechanisms of Alcohol Dehydration 202  
Mechanism 5.1 The  $E1$  Mechanism for Acid-Catalyzed Dehydration of *tert*-Butyl Alcohol 203
- 5.13 Rearrangements in Alcohol Dehydration 204  
Mechanism 5.2 Carbocation Rearrangement in Dehydration of 3,3-Dimethyl-2-butanol 205  
Mechanism 5.3 Hydride Shift in Dehydration of 1-Butanol 207
- 5.14 Dehydrohalogenation of Alkyl Halides 208
- 5.15 The  $E2$  Mechanism of Dehydrohalogenation of Alkyl Halides 210  
Mechanism 5.4  $E2$  Elimination of an Alkyl Halide 211
- 5.16 Anti Elimination in  $E2$  Reactions: Stereoelectronic Effects 212
- 5.17 Isotope Effects and the  $E2$  Mechanism 213
- 5.18 The  $E1$  Mechanism of Dehydrohalogenation of Alkyl Halides 214  
Mechanism 5.5 The  $E1$  Mechanism for Dehydrohalogenation of 2-Bromo-2-methylbutane in Ethanol 215
- 5.19 **Summary** 216  
**Problems** 220  
**Descriptive Passage and Interpretive Problems 5:**  
**A Mechanistic Preview of Addition Reactions** 224

## CHAPTER 6

### Addition Reactions of Alkenes 226

- 6.1 Hydrogenation of Alkenes 227
- 6.2 Heats of Hydrogenation 228  
Mechanism 6.1 Hydrogenation of Alkenes 229

- 6.3 Stereochemistry of Alkene Hydrogenation 230
- 6.4 Electrophilic Addition of Hydrogen Halides to Alkenes 232
- 6.5 Regioselectivity of Hydrogen Halide Addition: Markovnikov's Rule 233
  - Mechanism 6.2 Electrophilic Addition of a Hydrogen Halide to an Alkene 233
- 6.6 Mechanistic Basis for Markovnikov's Rule 235
  - Rules, Laws, Theories, and the Scientific Method 237
- 6.7 Carbocation Rearrangements in Hydrogen Halide Addition to Alkenes 237
- 6.8 Addition of Sulfuric Acid to Alkenes 239
- 6.9 Acid-Catalyzed Hydration of Alkenes 240
  - Mechanism 6.3 Acid-Catalyzed Hydration of 2-Methylpropene 241
- 6.10 Thermodynamics of Addition-Elimination Equilibria 242
- 6.11 Hydroboration-Oxidation of Alkenes 245
- 6.12 Stereochemistry of Hydroboration-Oxidation 247
- 6.13 Mechanism of Hydroboration-Oxidation 247
  - Mechanism 6.4 Hydroboration of 1-Methylcyclopentene 248
  - Mechanism 6.5 Oxidation of an Organoborane 249
- 6.14 Addition of Halogens to Alkenes 250
- 6.15 Stereochemistry of Halogen Addition 250
- 6.16 Mechanism of Halogen Addition to Alkenes: Halonium Ions 251
  - Mechanism 6.6 Electrophilic Addition of Bromine to Ethylene 252
  - Mechanism 6.7 Formation of Bromohydrin 253
- 6.17 Conversion of Alkenes to Vicinal Halohydrins 253
- 6.18 Free-Radical Addition of Hydrogen Bromide to Alkenes 254
  - Mechanism 6.8 Free-Radical Addition of Hydrogen Bromide to 1-Butene 256
- 6.19 Epoxidation of Alkenes 257
  - Mechanism 6.9 Epoxidation of an Alkene 259
- 6.20 Ozonolysis of Alkenes 259
- 6.21 Reactions of Alkenes with Alkenes: Polymerization 261
  - Mechanism 6.10 Acid-Catalyzed Dimerization of 2-Methylpropene 262
  - Ethylene and Propene: The Most Important Industrial Organic Chemicals 263
  - Mechanism 6.11 Free-Radical Polymerization of Ethylene 264
- 6.22 Summary 266
  - Problems 269
  - Descriptive Passage and Interpretive Problems 6: Oxymercuration 275

## 7

### CHAPTER

## Stereochemistry 278

- 7.1 Molecular Chirality: Enantiomers 279
- 7.2 The Chirality Center 281

- 7.3 Symmetry in Achiral Structures 283
- 7.4 Optical Activity 284
- 7.5 Absolute and Relative Configuration 286
- 7.6 The Cahn-Ingold-Prelog *R-S* Notational System 288
- 7.7 Fischer Projections 290
- 7.8 Properties of Enantiomers 292
- 7.9 Chirality Axis 293
  - Chiral Drugs 294
- 7.10 Reactions That Create a Chirality Center 296
- 7.11 Chiral Molecules with Two Chirality Centers 299
- 7.12 Achiral Molecules with Two Chirality Centers 301
- 7.13 Molecules with Multiple Chirality Centers 303
  - Chirality of Disubstituted Cyclohexanes 304
- 7.14 Reactions That Produce Diastereomers 305
- 7.15 Resolution of Enantiomers 307
- 7.16 Stereoregular Polymers 309
- 7.17 Chirality Centers Other Than Carbon 310
- 7.18 Summary 311
  - Problems 314
  - Descriptive Passage and Interpretive Problems 7: Prochirality 320

## 8

### CHAPTER

## Nucleophilic Substitution 322

- 8.1 Functional Group Transformation by Nucleophilic Substitution 323
- 8.2 Relative Reactivity of Halide Leaving Groups 326
- 8.3 The  $S_N2$  Mechanism of Nucleophilic Substitution 327
  - Mechanism 8.1 The  $S_N2$  Mechanism of Nucleophilic Substitution 327
- 8.4 Steric Effects in  $S_N2$  Reaction Rates 330
- 8.5 Nucleophiles and Nucleophilicity 332
- 8.6 The  $S_N1$  Mechanism of Nucleophilic Substitution 334
  - Enzyme-Catalyzed Nucleophilic Substitutions of Alkyl Halides 335
  - Mechanism 8.2 The  $S_N1$  Mechanism of Nucleophilic Substitution 336
- 8.7 Carbocation Stability and  $S_N1$  Reaction Rates 337
- 8.8 Stereochemistry of  $S_N1$  Reactions 338
- 8.9 Carbocation Rearrangements in  $S_N1$  Reactions 339
- 8.10 Effect of Solvent on the Rate of Nucleophilic Substitution 340
  - Mechanism 8.3 Carbocation Rearrangement in the  $S_N1$  Hydrolysis of 2-Bromo-3-methylbutane 340
- 8.11 Substitution and Elimination as Competing Reactions 344
- 8.12 Nucleophilic Substitution and Elimination of Alkyl Sulfonates 347

- 8.13 Summary 350  
 Problems 351  
 Descriptive Passage and Interpretive Problems 8:  
 Nucleophilic Substitution 356

## CHAPTER 9

### Alkynes 359

- 9.1 Sources of Alkynes 360  
 9.2 Nomenclature 362  
 9.3 Physical Properties of Alkynes 362  
 9.4 Structure and Bonding in Alkynes:  
 $sp$  Hybridization 362  
 9.5 Acidity of Acetylene and Terminal Alkynes 365  
 9.6 Preparation of Alkynes by Alkylation of Acetylene  
 and Terminal Alkynes 367  
 9.7 Preparation of Alkynes by Elimination Reactions 368  
 9.8 Reactions of Alkynes 370  
 9.9 Hydrogenation of Alkynes 370  
 9.10 Metal-Ammonia Reduction of Alkynes 372  
 9.11 Addition of Hydrogen Halides to Alkynes 373  
 Mechanism 9.1 Sodium-Ammonia Reduction of an  
 Alkyne 373  
 9.12 Hydration of Alkynes 375  
 Mechanism 9.2 Conversion of an Enol to a  
 Ketone 376  
 9.13 Addition of Halogens to Alkynes 377  
 Some Things Can Be Made from Acetylene . . .  
 But Aren't 378  
 9.14 Ozonolysis of Alkynes 378  
 9.15 Summary 379  
 Problems 382  
 Descriptive Passage and Interpretive Problems 9:  
 Thinking Mechanistically About Alkynes 386

## CHAPTER 10

### Conjugation in Alkadienes and Allylic Systems 388

- 10.1 The Allyl Group 389  
 10.2 Allylic Carbocations 390  
 10.3  $S_N1$  Reactions of Allylic Halides 392  
 Mechanism 10.1 Hydrolysis of an Allylic Halide 393  
 10.4  $S_N2$  Reactions of Allylic Halides 394  
 10.5 Allylic Free Radicals 395  
 10.6 Allylic Halogenation 396  
 Mechanism 10.2 Allylic Chlorination of Propene 397  
 10.7 Allylic Anions 399  
 10.8 Classes of Dienes 400  
 10.9 Relative Stabilities of Dienes 401

- 10.10 Bonding in Conjugated Dienes 402  
 10.11 Bonding in Allenes 404  
 10.12 Preparation of Dienes 405  
 Diene Polymers 406  
 10.13 Addition of Hydrogen Halides to Conjugated  
 Dienes 407  
 Mechanism 10.3 Addition of Hydrogen Chloride  
 to 1,3-Cyclopentadiene 408  
 10.14 Halogen Addition to Dienes 409  
 10.15 The Diels-Alder Reaction 410  
 10.16 The  $\pi$  Molecular Orbitals of Ethylene and  
 1,3-Butadiene 415  
 10.17 A  $\pi$  Molecular Orbital Analysis of the Diels-Alder  
 Reaction 417  
 Mechanism 10.4 Orbital Interaction in the Diels-Alder  
 Reaction 417  
 10.18 Summary 418  
 Problems 421  
 Descriptive Passage and Interpretive Problems 10:  
 Intramolecular and Retro Diels-Alder Reactions 425

## CHAPTER 11

### Arenes and Aromaticity 428

- 11.1 Benzene 429  
 11.2 The Structure of Benzene 430  
 11.3 The Stability of Benzene 432  
 11.4 An Orbital Hybridization View of Bonding in  
 Benzene 433  
 11.5 The  $\alpha$  Molecular Orbitals of Benzene 434  
 11.6 Substituted Derivatives of Benzene and Their  
 Nomenclature 435  
 11.7 Polycyclic Aromatic Hydrocarbons 438  
 11.8 Physical Properties of Arenes 439  
 Carbon Clusters, Fullerenes, and Nanotubes 440  
 11.9 Reactions of Arenes: A Preview 440  
 11.10 The Birch Reduction 442  
 11.11 Free-Radical Halogenation of Alkylbenzenes 442  
 Mechanism 11.1 The Birch Reduction 443  
 11.12 Oxidation of Alkylbenzenes 446  
 11.13  $S_N1$  Reactions of Benzylic Halides 448  
 11.14  $S_N2$  Reactions of Benzylic Halides 449  
 11.15 Preparation of Alkenylbenzenes 450  
 11.16 Addition Reactions of Alkenylbenzenes 451  
 11.17 Polymerization of Styrene 453  
 Mechanism 11.2 Free-Radical Polymerization of  
 Styrene 453  
 11.18 Cyclobutadiene and Cyclooctatetraene 454  
 11.19 Hückel's Rule 456  
 11.20 Annulenes 458  
 11.21 Aromatic Ions 460  
 11.22 Heterocyclic Aromatic Compounds 463  
 11.23 Heterocyclic Aromatic Compounds and Hückel's  
 Rule 465  
 11.24 Summary 467

Problems 470

Descriptive Passage and Interpretive Problems 11:

The Hammett Equation 474

## CHAPTER 12

### Reactions of Arenes: Electrophilic and Nucleophilic Aromatic Substitution 478

- 12.1 Representative Electrophilic Aromatic Substitution Reactions of Benzene 479
- 12.2 Mechanistic Principles of Electrophilic Aromatic Substitution 480
- 12.3 Nitration of Benzene 482
  - Mechanism 12.1 Nitration of Benzene 483
- 12.4 Sulfonation of Benzene 484
- 12.5 Halogenation of Benzene 484
  - Mechanism 12.2 Sulfonation of Benzene 485
  - Biosynthetic Halogenation 486
  - Mechanism 12.3 Bromination of Benzene 486
- 12.6 Friedel-Crafts Alkylation of Benzene 488
  - Mechanism 12.4 Friedel-Crafts Alkylation 489
- 12.7 Friedel-Crafts Acylation of Benzene 490
  - Mechanism 12.5 Friedel-Crafts Acylation 491
- 12.8 Synthesis of Alkylbenzenes by Acylation-Reduction 492
- 12.9 Rate and Regioselectivity in Electrophilic Aromatic Substitution 494
- 12.10 Rate and Regioselectivity in the Nitration of Toluene 495
- 12.11 Rate and Regioselectivity in the Nitration of (Trifluoromethyl) Benzene 497
- 12.12 Substituent Effects in Electrophilic Aromatic Substitution: Activating Substituents 499
- 12.13 Substituent Effects in Electrophilic Aromatic Substitution: Strongly Deactivating Substituents 503
- 12.14 Substituent Effects in Electrophilic Aromatic Substitution: Halogens 506
- 12.15 Multiple Substituent Effects 507
- 12.16 Regioselective Synthesis of Disubstituted Aromatic Compounds 510
- 12.17 Substitution in Naphthalene 512
- 12.18 Substitution in Heterocyclic Aromatic Compounds 513
- 12.19 Nucleophilic Aromatic Substitution 514
- 12.20 Nucleophilic Substitution in Nitro-Substituted Aryl Halides 515
- 12.21 The Addition-Elimination Mechanism of Nucleophilic Aromatic Substitution 516
  - Mechanism 12.6 Nucleophilic Aromatic Substitution in *p*-Fluoronitrobenzene by the Addition-Elimination Mechanism 518
- 12.22 Related Nucleophilic Aromatic Substitutions 520
- 12.23 Summary 521
  - Problems 525
  - Descriptive Passage and Interpretive Problems 12: Benzene 534

## CHAPTER 13

### Spectroscopy 538

- 13.1 Principles of Molecular Spectroscopy: Electromagnetic Radiation 539
- 13.2 Principles of Molecular Spectroscopy: Quantized Energy States 541
- 13.3 Introduction to  $^1\text{H}$  NMR Spectroscopy 541
- 13.4 Nuclear Shielding and  $^1\text{H}$  Chemical Shifts 543
- 13.5 Effects of Molecular Structure on  $^1\text{H}$  Chemical Shifts 546
  - Ring Currents—Aromatic and Antiaromatic 551
- 13.6 Interpreting  $^1\text{H}$  NMR Spectra 552
- 13.7 Spin-Spin Splitting in  $^1\text{H}$  NMR Spectroscopy 555
- 13.8 Splitting Patterns: The Ethyl Group 557
- 13.9 Splitting Patterns: The Isopropyl Group 559
- 13.10 Splitting Patterns: Pairs of Doublets 559
- 13.11 Complex Splitting Patterns 561
- 13.12  $^1\text{H}$  NMR Spectra of Alcohols 563
  - Magnetic Resonance Imaging (MRI) 564
- 13.13 NMR and Conformations 564
- 13.14  $^{13}\text{C}$  NMR Spectroscopy 565
- 13.15  $^{13}\text{C}$  Chemical Shifts 567
- 13.16  $^{13}\text{C}$  NMR and Peak Intensities 569
- 13.17  $^{13}\text{C}$ – $^1\text{H}$  Coupling 570
- 13.18 Using DEPT to Count Hydrogens Attached to  $^{13}\text{C}$  570
- 13.19 2D NMR: COSY and HETCOR 572
- 13.20 Introduction to Infrared Spectroscopy 574
  - Spectra by the Thousands 575
- 13.21 Infrared Spectra 576
- 13.22 Characteristic Absorption Frequencies 578
- 13.23 Ultraviolet-Visible (UV-VIS) Spectroscopy 582
- 13.24 Mass Spectrometry 584
- 13.25 Molecular Formula as a Clue to Structure 589
- 13.26 Summary 590
  - Problems 593
  - Descriptive Passage and Interpretive Problems 13: Calculating Aromatic  $^{13}\text{C}$  Chemical Shifts 603

## CHAPTER 14

### Organometallic Compounds 606

- 14.1 Organometallic Nomenclature 607
- 14.2 Carbon-Metal Bonds in Organometallic Compounds 608
- 14.3 Preparation of Organolithium Compounds 609
- 14.4 Preparation of Organomagnesium Compounds: Grignard Reagents 610
- 14.5 Organolithium and Organomagnesium Compounds as Brønsted Bases 612
- 14.6 Synthesis of Alcohols Using Grignard Reagents 614

- 14.7 Synthesis of Alcohols Using Organolithium Reagents 616
- 14.8 Synthesis of Acetylenic Alcohols 616
- 14.9 Retrosynthetic Analysis 617
- 14.10 Alkane Synthesis Using Organocopper Reagents 620
  - Mechanism 14.1 Formation of a Lithium Diaklycuprate (Gilman Reagent) 621
- 14.11 An Organozinc Reagent for Cyclopropane Synthesis 622
- 14.12 Carbenes and Carbenoids 623
  - Mechanism 14.2 Similarities Between the Mechanisms of Reaction of an Alkene with Iodomethylzinc Iodide and a Peroxy Acid 624
- 14.13 Transition-Metal Organometallic Compounds 625
  - An Organometallic That Occurs Naturally: Coenzyme B<sub>12</sub> 627
- 14.14 Homogeneous Catalytic Hydrogenation 628
  - Mechanism 14.3 Homogeneous Hydrogenation of Propene in the Presence of Wilkinson's Catalyst 629
- 14.15 Olefin Metathesis 631
  - Mechanism 14.4 Olefin Cross-Metathesis 632
- 14.16 Ziegler-Natta Catalysis of Alkene Polymerization 634
  - Mechanism 14.5 Polymerization of Ethylene in the Presence of a Ziegler-Natta Catalyst 635
- 14.17 Summary 636
  - Problems 639
  - Descriptive Passage and Interpretive Problems 14: The Heck Reaction 643

## CHAPTER 15

### Alcohols, Diols, and Thiols 646

- 15.1 Sources of Alcohols 647
- 15.2 Preparation of Alcohols by Reduction of Aldehydes and Ketones 648
  - Mechanism 15.1 Sodium Borohydride Reduction of an Aldehyde or Ketone 653
- 15.3 Preparation of Alcohols by Reduction of Carboxylic Acids 654
- 15.4 Preparation of Alcohols from Epoxides 654
- 15.5 Preparation of Diols 656
- 15.6 Reactions of Alcohols: A Review and a Preview 658
- 15.7 Conversion of Alcohols to Ethers 658
- 15.8 Esterification 660
  - Mechanism 15.2 Acid-Catalyzed Formation of Diethyl Ether from Ethyl Alcohol 660
- 15.9 Oxidation of Alcohols 663
  - Mechanism 15.3 Chromic Acid Oxidation of 2-Propanol 665
- 15.10 Biological Oxidation of Alcohols 666
  - Mechanism 15.4 Dimethyl Sulfoxide Oxidation of an Alcohol 666
  - Economic and Environmental Factors in Organic Synthesis 667
- 15.11 Oxidative Cleavage of Vicinal Diols 669
  - Mechanism 15.5 Oxidation of Ethanol by NAD<sup>+</sup> 669
- 15.12 Thiols 670
- 15.13 Spectroscopic Analysis of Alcohols and Thiols 674
- 15.14 Summary 675
  - Problems 679
  - Descriptive Passage and Interpretive Problems 15: The Pinacol Rearrangement 684

## CHAPTER 16

### Ethers, Epoxides, and Sulfides 686

- 16.1 Nomenclature of Ethers, Epoxides, and Sulfides 687
- 16.2 Structure and Bonding in Ethers and Epoxides 688
- 16.3 Physical Properties of Ethers 689
- 16.4 Crown Ethers 690
- 16.5 Preparation of Ethers 692
  - Polyether Antibiotics 693
- 16.6 The Williamson Ether Synthesis 694
- 16.7 Reactions of Ethers: A Review and a Preview 695
- 16.8 Acid-Catalyzed Cleavage of Ethers 696
  - Mechanism 16.1 Cleavage of Ethers by Hydrogen Halides 697
- 16.9 Preparation of Epoxides: A Review and a Preview 698
- 16.10 Conversion of Vicinal Halohydrins to Epoxides 699
- 16.11 Reactions of Epoxides: A Review and a Preview 700
- 16.12 Nucleophilic Ring Opening of Epoxides 701
  - Mechanism 16.2 Nucleophilic Ring Opening of an Epoxide 703
- 16.13 Acid-Catalyzed Ring Opening of Epoxides 703
  - Mechanism 16.3 Acid-Catalyzed Ring Opening of Ethylene Oxide 704
- 16.14 Epoxides in Biological Processes 706
- 16.15 Preparation of Sulfides 706
- 16.16 Oxidation of Sulfides: Sulfoxides and Sulfones 707
- 16.17 Alkylation of Sulfides: Sulfonium Salts 708
  - Mechanism 16.4 Nucleophilic Substitution of Adenosine Triphosphate (ATP) by Methionine 709
- 16.18 Spectroscopic Analysis of Ethers, Epoxides, and Sulfides 709
- 16.19 Summary 711
  - Problems 715
  - Descriptive Passage and Interpretive Problems 16: Epoxide Rearrangements and the NIH Shift 721

## CHAPTER 17

### Aldehydes and Ketones: Nucleophilic Addition to the Carbonyl Group 724

- 17.1 Nomenclature 725
- 17.2 Structure and Bonding: The Carbonyl Group 728
- 17.3 Physical Properties 730
- 17.4 Sources of Aldehydes and Ketones 730

- 17.5 Reactions of Aldehydes and Ketones: A Review and a Preview 734
- 17.6 Principles of Nucleophilic Addition: Hydration of Aldehydes and Ketones 735
  - Mechanism 17.1 Hydration of an Aldehyde or Ketone in Basic Solution 738
- 17.7 Cyanohydrin Formation 739
  - Mechanism 17.2 Hydration of an Aldehyde or Ketone in Acid Solution 739
  - Mechanism 17.3 Cyanohydrin Formation 740
- 17.8 Acetal Formation 742
  - Mechanism 17.4 Acetal Formation from Benzaldehyde and Ethanol 743
- 17.9 Acetals as Protecting Groups 745
- 17.10 Reaction with Primary Amines: Imines 746
  - Mechanism 17.5 Imine Formation from Benzaldehyde and Methylamine 747
  - Imines in Biological Chemistry 749
- 17.11 Reaction with Secondary Amines: Enamines 751
  - Mechanism 17.6 Enamine Formation from Cyclopentanone and Pyrrolidine 752
- 17.12 The Wittig Reaction 752
  - Mechanism 17.7 The Wittig Reaction 754
- 17.13 Planning an Alkene Synthesis via the Wittig Reaction 755
- 17.14 Stereoselective Addition to Carbonyl Groups 757
- 17.15 Oxidation of Aldehydes 758
- 17.16 Spectroscopic Analysis of Aldehydes and Ketones 759
- 17.17 Summary 761
  - Problems 764
  - Descriptive Passage and Interpretive Problems 17: The Baeyer-Villiger Oxidation 772

## CHAPTER 18

### Carboxylic Acids 776

- 18.1 Carboxylic Acid Nomenclature 777
- 18.2 Structure and Bonding 779
- 18.3 Physical Properties 780
- 18.4 Acidity of Carboxylic Acids 780
- 18.5 Substituents and Acid Strength 783
- 18.6 Ionization of Substituted Benzoic Acids 785
- 18.7 Salts of Carboxylic Acids 786
- 18.8 Dicarboxylic Acids 788
- 18.9 Carbonic Acid 789
- 18.10 Sources of Carboxylic Acids 790
- 18.11 Synthesis of Carboxylic Acids by the Carboxylation of Grignard Reagents 792
- 18.12 Synthesis of Carboxylic Acids by the Preparation and Hydrolysis of Nitriles 793
- 18.13 Reactions of Carboxylic Acids: A Review and a Preview 794
- 18.14 Mechanism of Acid-Catalyzed Esterification 794
  - Mechanism 18.1 Acid-Catalyzed Esterification of Benzoic Acid with Methanol 796

- 18.15 Intramolecular Ester Formation: Lactones 798
- 18.16 Decarboxylation of Malonic Acid and Related Compounds 799
- 18.17 Spectroscopic Properties of Carboxylic Acids 802
- 18.18 Summary 803
  - Problems 805
  - Descriptive Passage and Interpretive Problems 18: Lactonization Methods 809

## CHAPTER 19

### Carboxylic Acid Derivatives: Nucleophilic Acyl Substitution 812

- 19.1 Nomenclature of Carboxylic Acid Derivatives 814
- 19.2 Structure and Reactivity of Carboxylic Acid Derivatives 815
- 19.3 General Mechanism for Nucleophilic Acyl Substitution 818
- 19.4 Nucleophilic Acyl Substitution in Acyl Chlorides 820
  - Mechanism 19.1 Acid-Catalyzed Hydrolysis of an Acyl Chloride via a Tetrahedral Intermediate 822
- 19.5 Nucleophilic Acyl Substitution in Acid Anhydrides 823
  - Mechanism 19.2 Nucleophilic Acyl Substitution in an Anhydride 824
- 19.6 Sources of Esters 825
- 19.7 Physical Properties of Esters 827
- 19.8 Reactions of Esters: A Preview 827
- 19.9 Acid-Catalyzed Ester Hydrolysis 829
  - Mechanism 19.3 Acid-Catalyzed Ester Hydrolysis 830
- 19.10 Ester Hydrolysis in Base: Saponification 832
  - Mechanism 19.4 Ester Hydrolysis in Basic Solution 834
- 19.11 Reaction of Esters with Ammonia and Amines 835
- 19.12 Reaction of Esters with Grignard Reagents: Synthesis of Tertiary Alcohols 836
  - Mechanism 19.5 Reaction of an Ester with a Grignard Reagent 837
- 19.13 Reaction of Esters with Lithium Aluminum Hydride 838
- 19.14 Amides 839
- 19.15 Hydrolysis of Amides 843
  - Mechanism 19.6 Amide Hydrolysis in Acid Solution 844
  - Mechanism 19.7 Amide Hydrolysis in Basic Solution 846
- 19.16 Lactams 847
  - $\beta$ -Lactam Antibiotics 847
- 19.17 Preparation of Nitriles 848
- 19.18 Hydrolysis of Nitriles 849
- 19.19 Addition of Grignard Reagents to Nitriles 850
  - Mechanism 19.8 Nitrile Hydrolysis in Basic Solution 851
- 19.20 Spectroscopic Analysis of Carboxylic Acid Derivatives 852

- 19.21 Summary 853  
 Problems 856  
 Descriptive Passage and Interpretive Problems 20:  
 Thioesters 863

## CHAPTER 20

### Enols and Enolates 866

- 20.1 Aldehyde, Ketone, and Ester Enolates 867  
 20.2 Enolate Regiochemistry 872  
 20.3 The Aldol Condensation 873  
   Mechanism 20.1 Aldol Addition of Butanal 874  
   Mechanism 20.2 Dehydration in a Base-Catalyzed Aldol Condensation 876  
 20.4 Mixed Aldol Condensations 878  
   Chalcones: From the Mulberry Tree to Cancer Chemotherapy 880  
 20.5 The Claisen Condensation 882  
   Mechanism 20.3 The Claisen Condensation of Ethyl Acetate 883  
 20.6 Intramolecular Claisen Condensation: The Dieckmann Cyclization 884  
 20.7 Mixed Claisen Condensations 885  
 20.8 Acylation of Ketones with Esters 886  
 20.9 Alkylation of Enolates 887  
 20.10 The Acetoacetic Ester Synthesis 889  
 20.11 The Malonic Ester Synthesis 891  
 20.12 Alkylation of Chiral Enolates 893  
 20.13 Enolization and Enol Content 895  
   Mechanism 20.4 Base-Catalyzed Enolization of an Aldehyde or Ketone in Aqueous Solution 899  
   Mechanism 20.5 Acid-Catalyzed Enolization of an Aldehyde or Ketone in Aqueous Solution 899  
 20.14  $\alpha$  Halogenation of Aldehydes and Ketones 900  
   Mechanism 20.6 Acid-Catalyzed Bromination of Acetone 901  
   Mechanism 20.7 Cleavage of a Tribromomethyl Ketone 903  
 20.15  $\alpha$  Halogenation of Carboxylic Acids: The Hell-Volhard-Zelinsky Reaction 904  
   The Haloform Reaction and the Biosynthesis of Trihalomethanes 904  
 20.16 Some Chemical and Stereochemical Consequences of Enolization 906  
 20.17 Effects of Conjugation in  $\alpha,\beta$ -Unsaturated Aldehydes and Ketones 907  
 20.18 Conjugate Addition to  $\alpha,\beta$ -Unsaturated Carbonyl Compounds 908  
 20.19 Addition of Carbanions to  $\alpha,\beta$ -Unsaturated Ketones: The Michael Reaction 910  
 20.20 Conjugate Addition of Organocopper Reagents to  $\alpha,\beta$ -Unsaturated Carbonyl Compounds 912

- 20.21 Summary 913  
 Problems 917  
 Descriptive Passage and Interpretive Problems 20:  
 The Enolate Chemistry of Dianions 926

## CHAPTER 21

### Amines 930

- 21.1 Amine Nomenclature 931  
 21.2 Structure and Bonding 933  
 21.3 Physical Properties 935  
 21.4 Basicity of Amines 936  
   Amines as Natural Products 941  
 21.5 Tetraalkylammonium Salts as Phase-Transfer Catalysts 942  
 21.6 Reactions That Lead to Amines: A Review and a Preview 943  
 21.7 Preparation of Amines by Alkylation of Ammonia 945  
 21.8 The Gabriel Synthesis of Primary Alkylamines 946  
 21.9 Preparation of Amines by Reduction 947  
   Mechanism 21.1 Lithium Aluminum Hydride Reduction of an Amide 950  
 21.10 Reductive Amination 951  
 21.11 Reactions of Amines: A Review and a Preview 952  
 21.12 Reaction of Amines with Alkyl Halides 954  
 21.13 The Hofmann Elimination 954  
 21.14 Electrophilic Aromatic Substitution in Arylamines 956  
 21.15 Nitrosation of Alkylamines 958  
 21.16 Nitrosation of Arylamines 960  
   Mechanism 21.2 Reactions of an Alkyl Diazonium Ion 960  
 21.17 Synthetic Transformations of Aryl Diazonium Salts 961  
 21.18 Azo Coupling 965  
   From Dyes to Sulfa Drugs 966  
 21.19 Spectroscopic Analysis of Amines 967  
 21.20 Summary 970  
 Problems 976  
 Descriptive Passage and Interpretive Problems 21:  
 Synthetic Applications of Enamines 984

## CHAPTER 22

### Phenols 988

- 22.1 Nomenclature 989  
 22.2 Structure and Bonding 990  
 22.3 Physical Properties 991  
 22.4 Acidity of Phenols 992  
 22.5 Substituent Effects on the Acidity of Phenols 993  
 22.6 Sources of Phenols 995  
 22.7 Naturally Occurring Phenols 996  
 22.8 Reactions of Phenols: Electrophilic Aromatic Substitution 997

- 22.9 Acylation of Phenols 999
- 22.10 Carboxylation of Phenols: Aspirin and the Kolbe-Schmitt Reaction 1001
- 22.11 Preparation of Aryl Ethers 1002
  - James Bond, Oxidative Stress, and Antioxidant Phenols 1004
- 22.12 Cleavage of Aryl Ethers by Hydrogen Halides 1006
- 22.13 Claisen Rearrangement of Allyl Aryl Ethers 1007
- 22.14 Oxidation of Phenols: Quinones 1008
- 22.15 Spectroscopic Analysis of Phenols 1009
- 22.16 Summary 1011
  - Problems 1013
  - Descriptive Passage and Interpretive Problems 22: Directed Metalation of Aryl Ethers 1019

## CHAPTER 23

### Carbohydrates 1022

- 23.1 Classification of Carbohydrates 1023
- 23.2 Fischer Projections and D,L Notation 1024
- 23.3 The Aldotetroses 1025
- 23.4 Aldopentoses and Aldohexoses 1026
- 23.5 A Mnemonic for Carbohydrate Configurations 1028
- 23.6 Cyclic Forms of Carbohydrates: Furanose Forms 1029
- 23.7 Cyclic forms of Carbohydrates: Pyranose Forms 1032
- 23.8 Mutarotation 1035
  - Mechanism 23.1 Acid-Catalyzed Mutarotation of D-Glucopyranose 1037
- 23.9 Carbohydrate Conformation: The Anomeric Effect 1038
- 23.10 Ketoses 1039
- 23.11 Deoxy Sugars 1040
- 23.12 Amino Sugars 1041
- 23.13 Branched-Chain Carbohydrates 1042
- 23.14 Glycosides: The Fischer Glycosidation 1043
  - Mechanism 23.2 Preparation of Methyl D-Glucopyranosides by Fischer Glycosidation 1044
- 23.15 Disaccharides 1046
- 23.16 Polysaccharides 1048
  - How Sweet It Is! 1049
- 23.17 Reactions of Carbohydrates 1050
- 23.18 Reduction of Monosaccharides 1050
- 23.19 Oxidation of Monosaccharides 1051
- 23.20 Periodic Acid Oxidation 1053
- 23.21 Cyanohydrin Formation and Chain Extension 1054
- 23.22 Epimerization, Isomerization, and Retro-Aldol Cleavage 1055
- 23.23 Acylation and Alkylation of Carbohydrate Hydroxyl Groups 1056
- 23.24 Glycosides: Synthesis of Oligosaccharides 1058
  - Mechanism 23.3 Silver-Assisted Glycosidation 1060
- 23.25 Glycobiology 1062
- 23.26 Summary 1064
  - Problems 1067
  - Descriptive Passage and Interpretive Problems 23: Emil Fischer and the Structure of (+)-Glucose 1072

## CHAPTER 24

### Lipids 1074

- 24.1 Acetyl Coenzyme A 1075
- 24.2 Fats, Oils, and Fatty Acids 1077
- 24.3 Fatty Acid Biosynthesis 1080
- 24.4 Phospholipids 1082
  - Mechanism 24.1 Biosynthesis of a Butanoyl Group from Acetyl and Malonyl Building Blocks 1082
- 24.5 Waxes 1085
- 24.6 Prostaglandins 1086
  - Nonsteroidal Anti-Inflammatory Drugs (NSAIDs) and COX-2 Inhibitors 1088
- 24.7 Terpenes: The Isoprene Rule 1090
- 24.8 Isopentenyl Pyrophosphate: The Biological Isoprene Unit 1093
- 24.9 Carbon-Carbon Bond Formation in Terpene Biosynthesis 1093
- 24.10 The Pathway from Acetate to Isopentenyl Diphosphate 1096
- 24.11 Steroids: Cholesterol 1098
  - Mechanism 24.2 Biosynthesis of Cholesterol from Squalene 1100
- 24.12 Vitamin D 1101
  - Good Cholesterol? Bad Cholesterol? What's the Difference? 1102
- 24.13 Bile Acids 1103
- 24.14 Corticosteroids 1103
- 24.15 Sex Hormones 1103
- 24.16 Carotenoids 1104
  - Crocuses Make Saffron from Carotenoids 1105
- 24.17 Summary 1106
  - Problems 1108
  - Descriptive Passage and Interpretive Problems 24: Polyketides 1112

## CHAPTER 25

### Amino Acids, Peptides, and Proteins 1116

- 25.1 Classification of Amino Acids 1118
- 25.2 Stereochemistry of Amino Acids 1123
- 25.3 Acid-Base Behavior of Amino Acids 1124
  - Electrophoresis 1127
- 25.4 Synthesis of Amino Acids 1128
- 25.5 Reactions of Amino Acids 1130
- 25.6 Some Biochemical Reactions of Amino Acids 1130
  - Mechanism 25.1 Pyridoxal 5'-Phosphate-Mediated Decarboxylation of an  $\alpha$ -Amino Acid 1131
  - Mechanism 25.2 Transamination: Biosynthesis of L-Alanine from L-Glutamic Acid and Pyruvic Acid 1135
- 25.7 Peptides 1137
- 25.8 Introduction to Peptide Structure Determination 1140

- 25.9 Amino Acid Analysis 1140
- 25.10 Partial Hydrolysis of Peptides 1141
- 25.11 End Group Analysis 1141
- 25.12 Insulin 1143
- 25.13 The Edman Degradation and Automated Sequencing of Peptides 1144
  - Mechanism 25.3 The Edman Degradation 1145
  - Peptide Mapping and MALDI Mass Spectrometry 1146
- 25.14 The Strategy of Peptide Synthesis 1147
- 25.15 Amino Group Protection 1148
- 25.16 Carboxyl Group Protection 1151
- 25.17 Peptide Bond Formation 1151
  - Mechanism 25.4 Amide Bond Formation Between a Carboxylic Acid and Amine Using N,N'-Dicyclohexylcarbodiimide 1152
- 25.18 Solid-Phase Peptide Synthesis: The Merrifield Method 1153
- 25.19 Secondary Structures of Polypeptides and Proteins 1155
- 25.20 Tertiary Structure of Peptides and Proteins 1159
  - Mechanism 25.5 Carboxypeptidase-Catalyzed Hydrolysis 1162
- 25.21 Coenzymes 1163
  - Oh NO! It's Inorganic! 1164
- 25.22 Protein Quaternary Structure: Hemoglobin 1164
- 25.23 G-Coupled Protein Receptors 1165
- 25.24 Summary 1166
  - Problems 1168
  - Descriptive Passage and Interpretive Problems 25: Amino Acids in Enantioselective Synthesis 1171

## CHAPTER 26

### Nucleosides, Nucleotides, and Nucleic Acids 1174

- 26.1 Pyrimidines and Purines 1175
- 26.2 Nucleosides 1178
- 26.3 Nucleotides 1180
- 26.4 Bioenergetics 1182
- 26.5 ATP and Bioenergetics 1182
- 26.6 Phosphodiester, Oligonucleotides, and Polynucleotides 1184
- 26.7 Nucleic Acids 1185
- 26.8 Secondary Structure of DNA: The Double Helix 1186
  - "It has not escaped our notice . . ." 1188
- 26.9 Tertiary Structure of DNA: Supercoils 1190
- 26.10 Replication of DNA 1191

- 26.11 Ribonucleic Acids 1193
- 26.12 Protein Biosynthesis 1196
- 26.13 AIDS 1197
- 26.14 DNA Sequencing 1198
- 26.15 The Human Genome Project 1200
- 26.16 DNA Profiling and the Polymerase Chain Reaction 1201
- 26.17 Recombinant DNA Technology 1204
- 26.18 Summary 1205
  - Problems 1208
  - Descriptive Passage and Interpretive Problems 26: Oligonucleotide Synthesis 1210

## CHAPTER 27

### Synthetic Polymers 1216

- 27.1 Some Background 1217
- 27.2 Polymer Nomenclature 1218
- 27.3 Classification of Polymers: Reaction Type 1219
- 27.4 Classification of Polymers: Chain Growth and Step Growth 1220
- 27.5 Classification of Polymers: Structure 1221
- 27.6 Classification of Polymers: Properties 1223
- 27.7 Addition Polymers: A Review and a Preview 1225
- 27.8 Chain Branching in Free-Radical Polymerization 1227
  - Mechanism 27.1 Branching in Polyethylene Caused by Intramolecular Hydrogen Transfer 1228
  - Mechanism 27.2 Branching in Polyethylene Caused by Intermolecular Hydrogen Transfer 1229
- 27.9 Anionic Polymerization: Living Polymers 1230
  - Mechanism 27.3 Anionic Polymerization of Styrene 1230
- 27.10 Cationic Polymerization 1232
- 27.11 Polyamides 1233
  - Mechanism 27.4 Cationic Polymerization of 2-Methylpropene 1233
- 27.12 Polyesters 1234
- 27.13 Polycarbonates 1236
- 27.14 Polyurethanes 1236
- 27.15 Copolymers 1237
  - Conducting Polymers 1239
- 27.16 Summary 1241
  - Problems 1243
  - Descriptive Passage and Interpretive Problems 27: Chemical Modification of Polymers 1245

Glossary G-1

Credits C-1

Index I-1

# 25 Amino Acids, Peptides, and Proteins

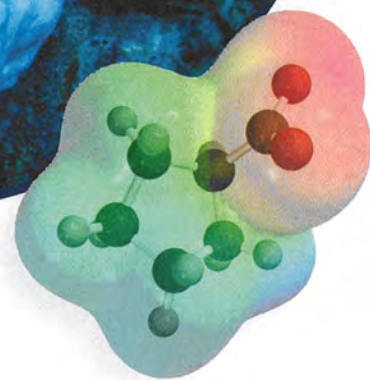
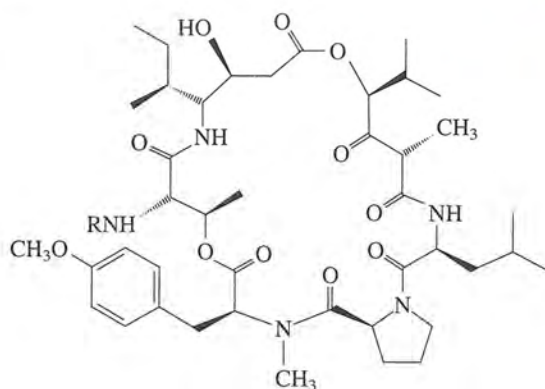
## Chapter Outline

- 25.1 Classification of Amino Acids 1118
- 25.2 Stereochemistry of Amino Acids 1123
- 25.3 Acid–Base Behavior of Amino Acids 1124
  - Electrophoresis 1127
- 25.4 Synthesis of Amino Acids 1128
- 25.5 Reactions of Amino Acids 1130
- 25.6 Some Biochemical Reactions of Amino Acids 1130
- 25.7 Peptides 1137
- 25.8 Introduction to Peptide Structure Determination 1140
- 25.9 Amino Acid Analysis 1140
- 25.10 Partial Hydrolysis of Peptides 1141
- 25.11 End Group Analysis 1141
- 25.12 Insulin 1143
- 25.13 The Edman Degradation and Automated Sequencing of Peptides 1144
  - Peptide Mapping and MALDI Mass Spectrometry 1146
- 25.14 The Strategy of Peptide Synthesis 1147
- 25.15 Amino Group Protection 1148
- 25.16 Carboxyl Group Protection 1151
- 25.17 Peptide Bond Formation 1151
- 25.18 Solid-Phase Peptide Synthesis: The Merrifield Method 1153
- 25.19 Secondary Structures of Peptides and Proteins 1155
- 25.20 Tertiary Structure of Polypeptides and Proteins 1159
- 25.21 Coenzymes 1163
  - Oh NO! It's Inorganic! 1164
- 25.22 Protein Quaternary Structure: Hemoglobin 1164
- 25.23 G-Coupled Protein Receptors 1165
- 25.24 Summary 1166
- Problems 1168
- Descriptive Passage and Interpretive Problems 25: Amino Acids in Enantioselective Synthesis 1171

## Mechanisms

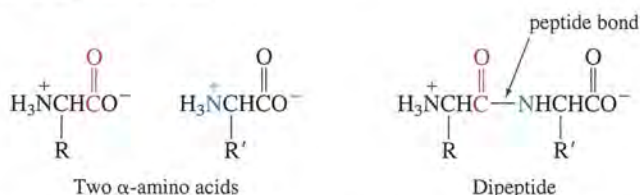
- 25.1 Pyridoxal 5'-Phosphate-Mediated Decarboxylation of an  $\alpha$ -Amino Acid 1131
- 25.2 Transamination: Biosynthesis of L-Alanine from L-Glutamic Acid and Pyruvic Acid 1135
- 25.3 The Edman Degradation 1145
- 25.4 Amide Bond Formation Between a Carboxylic Acid and an Amine Using *N,N'*-Dicyclohexylcarbodiimide 1152
- 25.5 Carboxypeptidase-Catalyzed Hydrolysis 1162

The didemnins are cyclic molecules composed of amino acids that were isolated from the marine invertebrate *Trididemnum solidum*. One of the amino acids is proline. The didemnins exhibit antitumor, antiviral, and immunosuppressive activity. The group labeled "R" in the structure is a side chain that contains additional amino acids.



**THE RELATIONSHIP** between structure and function reaches its ultimate expression in the chemistry of amino acids, peptides, and proteins.

Amino acids are carboxylic acids that contain an amine function. An amide bond between the carboxylic acid function of one amino acid and the amino nitrogen of another is called a **peptide bond**.



A **dipeptide** is a molecule consisting of two amino acids joined by a peptide bond. A **tripeptide** has three amino acids joined by two peptide bonds, a **tetrapeptide** has four amino acids, and so on. Peptides with more than 30–50 amino acids are **polypeptides**. **Proteins** are polypeptides that have some biological function.

The most striking thing about proteins is the diversity of their roles in living systems: silk is a protein, skin and hair are mostly proteins, many hormones are proteins, a protein carries oxygen from the lungs to the tissues where it is stored by another protein, and all enzymes are proteins.

As in most aspects of chemistry and biochemistry, structure is the key to function. We'll explore the structure of proteins by first concentrating on their fundamental building block units, the  $\alpha$ -amino acids. Then, after developing the principles of peptide structure, we'll see how the insights gained from these smaller molecules aid our understanding of proteins.

## 25.1 Classification of Amino Acids

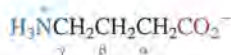
Amino acids are classified as  $\alpha$ ,  $\beta$ ,  $\gamma$ , and so on, according to the location of the amine group on the carbon chain that contains the carboxylic acid function.



1-Aminocyclopropanecarboxylic acid:  
an  $\alpha$ -amino acid that is the biological  
precursor to ethylene in plants

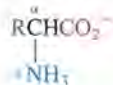


3-Aminopropanoic acid: known as  $\beta$ -alanine,  
it is a  $\beta$ -amino acid that makes up one of  
the structural units of coenzyme A

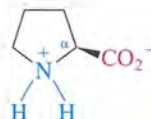


4-Aminobutanoic acid: known as  
 $\gamma$ -aminobutyric acid (GABA), it is a  $\gamma$ -amino  
acid and is involved in the transmission of  
nerve impulses

Although more than 700 different amino acids are known to occur naturally, the group of 20 called the **standard amino acids** listed in Table 25.1 (pages 1120–1121) commands special attention. These 20 are the amino acids coded for in DNA-directed protein synthesis. All are  $\alpha$ -amino acids, and all but one contain a primary amino function.



The one exception is proline, a secondary amine in which the amino nitrogen is incorporated into a five-membered ring.

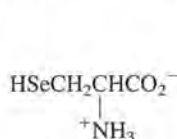


Proline

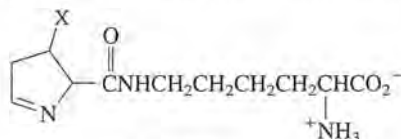
Table 25.1 includes three-letter and one-letter abbreviations for the amino acids. Both enjoy wide use.

Our bodies make some of the amino acids shown in the table. The others, which are called **essential amino acids**, we have to obtain from our diet.

When a protein contains an amino acid different from those in the table, it normally is formed by modification of one of the 20 rather than being coded for in DNA. Two exceptions have recently been discovered, selenocysteine (1986) and pyrrolysine (2002). They have been referred to as the “twenty-first and twenty-second amino acids,” respectively.



Selenocysteine



Pyrrolysine  
(X is uncertain; CH<sub>3</sub>, NH<sub>2</sub>, or OH)

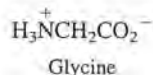
The most important aspect of Table 25.1 is that while the 20 amino acids share the common feature of being  $\alpha$ -amino acids, their side chains differ in respect to their:

1. Size and shape
2. Electronic characteristics, acid–base properties, and ability to engage in ionic bonding, covalent bonding, hydrogen bonding, and van der Waals forces.

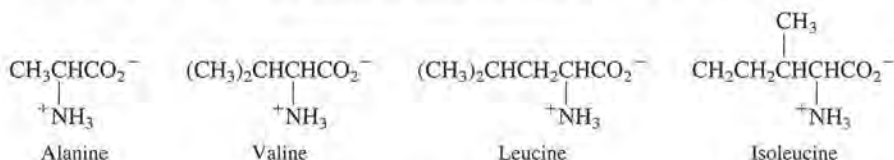
Table 25.1 shows the amino acids in the form in which they exist at a pH of 7: amine groups as positively charged ammonium ions, and carboxylic acid groups as negatively charged carboxylates. The electrostatic potential is mapped on the van

der Waals surface of the molecule and so displays the charge distribution, size, and shape at the same time.

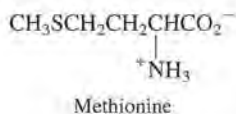
**Nonpolar Side Chains:** *Glycine* is the smallest amino acid because it has no side chain. The main service it offers is to the polypeptide chain itself. It can add length and flexibility to a polypeptide without sacrificing strength or making spatial demands of its own.



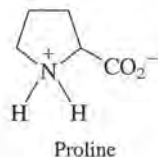
After glycine, the next four amino acids in the figure all have alkyl groups (R) as side chains: *alanine* (R = methyl), *valine* (R = isopropyl), *leucine* (R = isobutyl), and *isoleucine* (R = *sec*-butyl). All are hydrophobic side chains and although electronically similar, they differ in size. Alanine is slightly larger than glycine, valine slightly larger than alanine, leucine slightly larger than valine, and isoleucine somewhat more spherical than leucine.



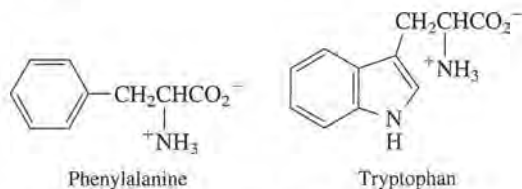
Compared with these, the presence of sulfur in its side chain makes *methionine* somewhat more polarizable and increases its ability to participate in dispersion forces.



*Proline* is relatively compact and has limited conformational flexibility because its side chain is cyclic. Moreover, amides of proline lack N—H bonds so cannot form hydrogen bonds. Consequently, the presence of proline affects the shape of a peptide chain more than most other amino acids.

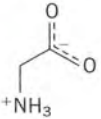
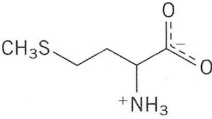
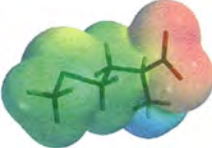
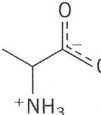
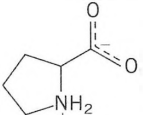
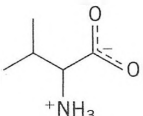
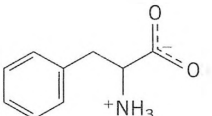
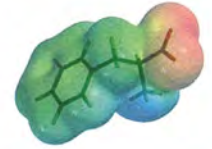
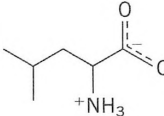
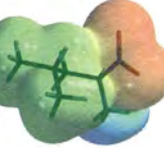
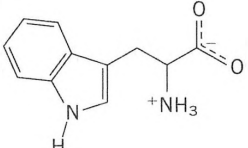
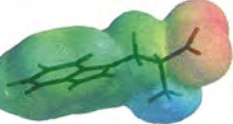
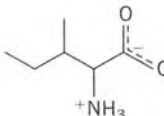
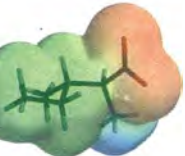
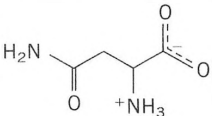
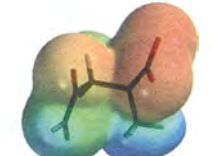


*Phenylalanine* and *tryptophan* have side chains that incorporate aromatic rings, which are large and hydrophobic. In addition to being larger than phenylalanine, tryptophan has a more electron-rich aromatic ring and is more polarizable. Its role is more specialized, and it is less abundant in proteins than most of the other amino acids.



**Amino Acids with Polar but Nonionized Side Chains:** Among amino acids with polar side chains, *serine* is the smallest; it is not much larger than alanine. With a —CH<sub>2</sub>OH side chain, serine participates well in hydrogen bonding and often occurs in regions of a peptide that are exposed to water. *Threonine* has a methyl group in place of one of the hydrogens of the —CH<sub>2</sub>OH group of serine, sterically hindering the OH group and making it less effective in hydrogen bonding.

TABLE 25.1 The Standard Amino Acids

| Name and abbreviation                        | Structural formula*                                                                 | Electrostatic potential map                                                         | Name and abbreviation                                    | Structural formula*                                                                   | Electrostatic potential map                                                           |
|----------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| <b>Amino acids with nonpolar side chains</b> |                                                                                     |                                                                                     |                                                          |                                                                                       |                                                                                       |
| Glycine<br>Gly (G)                           |    |    | Methionine†<br>Met (M)                                   |    |    |
| Alanine<br>Ala (A)                           |    |    | Proline<br>Pro (P)                                       |    |    |
| Valine†<br>Val (V)                           |    |    | Phenylalanine†<br>Phe (F)                                |    |    |
| Leucine†<br>Leu (L)                          |   |   | Tryptophan†<br>Trp (W)                                   |   |   |
| Isoleucine†<br>Ile (I)                       |  |  | <b>Amino acids with polar but nonionized side chains</b> |                                                                                       |                                                                                       |
|                                              |                                                                                     |                                                                                     | Asparagine<br>Asn (N)                                    |  |  |

\*All amino acids are shown in the form present in greatest concentration at pH 7.

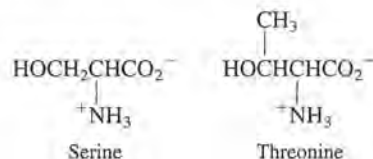
†An essential amino acid, which must be present in the diet of animals to ensure normal growth.

**TABLE 25.1 The Standard Amino Acids (Continued)**

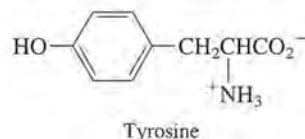
| Name and abbreviation                                                | Structural formula* | Electrostatic potential map | Name and abbreviation                      | Structural formula* | Electrostatic potential map |
|----------------------------------------------------------------------|---------------------|-----------------------------|--------------------------------------------|---------------------|-----------------------------|
| <b>Amino acids with polar but nonionized side chains (continued)</b> |                     |                             | <b>Amino acids with acidic side chains</b> |                     |                             |
| Glutamine<br>Gln (N)                                                 |                     |                             | Aspartic acid<br>Asp (D)                   |                     |                             |
| Serine<br>Ser (S)                                                    |                     |                             | Glutamic acid<br>Glu (E)                   |                     |                             |
| Threonine<br>Thr (T)                                                 |                     |                             | <b>Amino acids with basic side chains</b>  |                     |                             |
| Tyrosine<br>Tyr (Y)                                                  |                     |                             | Lysine<br>Lys (K)                          |                     |                             |
| Cysteine<br>Cys (C)                                                  |                     |                             | Arginine†<br>Arg (R)                       |                     |                             |
|                                                                      |                     |                             | Histidine†<br>His (H)                      |                     |                             |

\*All amino acids are shown in the form present in greatest concentration at pH 7.

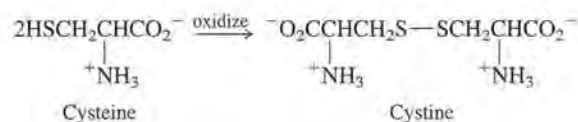
†An essential amino acid, which must be present in the diet of animals to ensure normal growth.



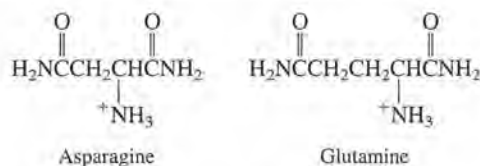
As a *p*-hydroxy derivative of phenylalanine, *tyrosine* has properties similar to it plus the ability to form hydrogen bonds involving its phenolic —OH group.



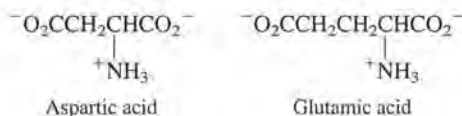
*Cysteine* is related to serine except that its side chain is —CH<sub>2</sub>SH rather than —CH<sub>2</sub>OH. The number of cysteines in a protein is often relatively small, but their effect on its three-dimensional shape is substantial. Oxidation of two cysteines converts their —CH<sub>2</sub>SH side chains to a —CH<sub>2</sub>S—SCH<sub>2</sub>— disulfide bridge between them (see Section 15.12). This ties together two often remote amino acids and helps guide the folding of the protein.



*Asparagine* and *glutamine* are amides. The side chains of both terminate in —C(=O)NH<sub>2</sub> and differ by only a single CH<sub>2</sub> group. Amide functions are quite polar and interact strongly with water molecules by hydrogen bonding. Like serine, asparagine and glutamine are often found in regions of a peptide that are in contact with water.

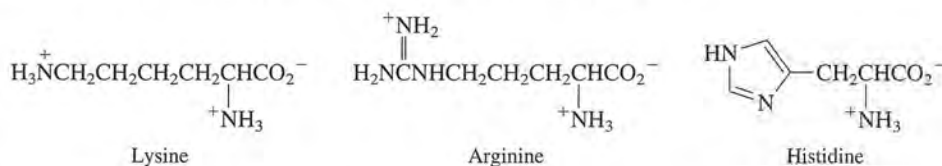


**Amino Acids with Acidic Side Chains:** The electrostatic potential maps of *aspartic acid* and *glutamic acid* are two of the most prominent ones in Table 25.1. The —CO<sub>2</sub>H side chains of both aspartic and glutamic acid are almost completely deprotonated to —CO<sub>2</sub><sup>−</sup> at biological pH, giving these species the most electron-rich units of all of the common amino acids. Their most important function is in ionic bonding to positively charged species. These ionic bonds can be to metal ions and —N<sup>+</sup> species among others.



**Amino Acids with Basic Side Chains:** Basic amino acids are the opposite of acidic amino acids. Their most important role is to form ionic bonds to negative ions—phosphate and the like. Lysine is a simple example. The side chain contains four CH<sub>2</sub> groups and terminates in —NH<sub>3</sub><sup>+</sup>. Arginine has an even more basic, if somewhat more complicated and larger, side chain. Conversely, the side chain of histidine is not as basic as that of

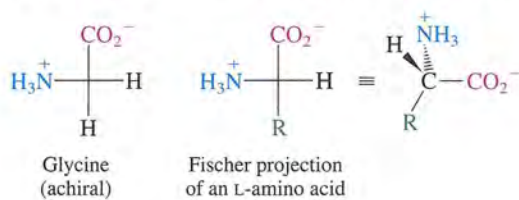
lysine and the concentrations of the unprotonated and protonated forms of histidine are almost equal at biological pH. Strong Lewis acid/Lewis base complexes between the unprotonated form of histidine and metal ions is very common in proteins. Histidine side chains are also involved in moving protons from one atom to another.



What is remarkable is not only the range of properties that are covered with just 20 amino acids, but also the fine tuning with respect to a particular property.

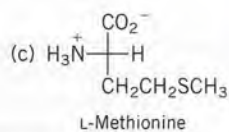
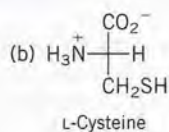
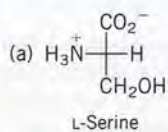
## 25.2 Stereochemistry of Amino Acids

Glycine is the only amino acid in Table 25.1 that is achiral. The  $\alpha$ -carbon atom is a chirality center in all the others. Configurations in amino acids are normally specified by the D, L notational system. All the chiral amino acids obtained from proteins have the L configuration at their  $\alpha$ -carbon atom, meaning that the amine group is at the left when a Fischer projection is arranged so the carboxyl group is at the top.

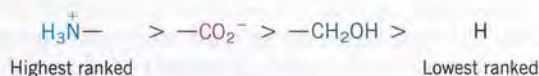


### Problem 25.1

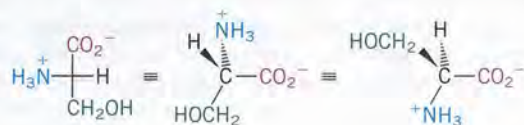
What is the absolute configuration (R or S) at the  $\alpha$ -carbon atom in each of the following L-amino acids?



**Sample Solution** (a) First identify the four groups attached directly to the chirality center, and rank them in order of decreasing sequence rule precedence. For L-serine these groups are

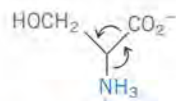


Next, translate the Fischer projection of L-serine to a three-dimensional representation, and orient it so that the lowest ranked substituent at the chirality center is directed away from you.



*Continued*

In order of decreasing precedence the three highest ranked groups trace a counterclockwise path.



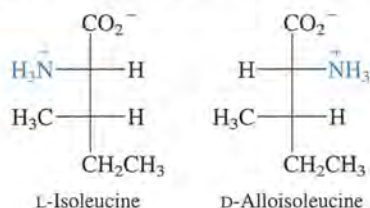
The absolute configuration of L-serine is S.

### Problem 25.2

The amino acid L-threonine is (2S,3R)-2-amino-3-hydroxybutanoic acid. Draw a Fischer projection for L-threonine.

Although all the chiral amino acids obtained from proteins have the L configuration at their  $\alpha$  carbon, that should not be taken to mean that D-amino acids are unknown. In fact, quite a number of D-amino acids occur naturally. D-Alanine, for example, is a constituent of bacterial cell walls and D-serine occurs in brain tissue. The point is that D-amino acids are not coded for by DNA.

A novel technique for dating archaeological samples called **amino acid racemization (AAR)** is based on the stereochemistry of amino acids. Over time, the configuration at the  $\alpha$ -carbon atom of a protein's amino acids is lost in a reaction that follows first-order kinetics. When the  $\alpha$  carbon is the only chirality center, this process corresponds to racemization. For an amino acid with two chirality centers, changing the configuration of the  $\alpha$  carbon from L to D gives a diastereomer. In the case of isoleucine, for example, the diastereomer is an amino acid not normally present in proteins, called *alloisoleucine*.

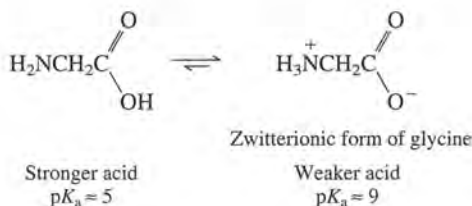


By measuring the L-isoleucine/D-alloisoleucine ratio in the protein isolated from the eggshells of an extinct Australian bird, a team of scientists recently determined that this bird lived approximately 50,000 years ago. Radiocarbon ( $^{14}\text{C}$ ) dating is not accurate for samples older than about 35,000 years, so AAR is a useful addition to the tools available to paleontologists.

## 25.3 Acid–Base Behavior of Amino Acids

The physical properties of a typical amino acid such as glycine suggest that it is a very polar substance, much more polar than would be expected on the basis of its formulation as  $\text{H}_2\text{NCH}_2\text{CO}_2\text{H}$ . Glycine is a crystalline solid; it does not melt, but on being heated it eventually decomposes at  $233^\circ\text{C}$ . It is very soluble in water but practically insoluble in nonpolar organic solvents. These properties are attributed to the fact that the stable form of glycine is a **zwitterion**.

The zwitterion is also often referred to as a *dipolar ion*. Note, however, that it is not an ion, but a neutral molecule.



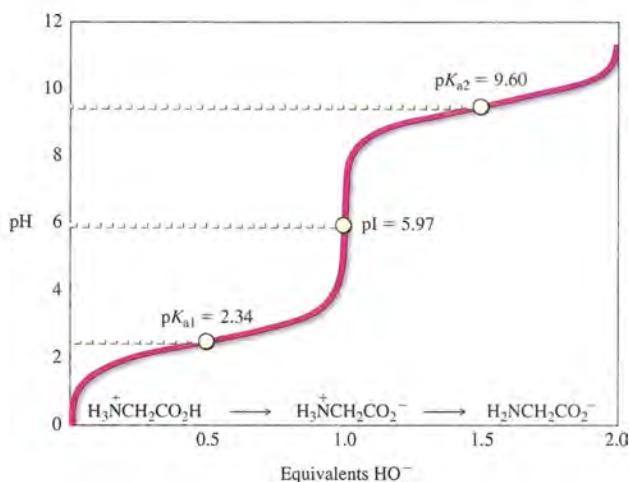
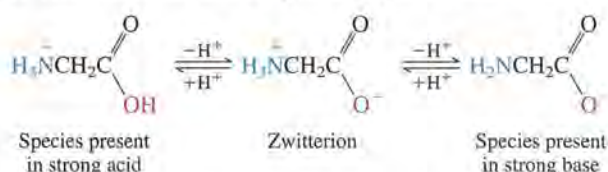


Figure 25.1

Titration curve of glycine. At pH values lower than 2.34,  $\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2\text{H}$  is the major species. At pH = 2.34,  $[\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2\text{H}] = [\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2^-]$ . Between pH = 2.34 and 9.60,  $\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2^-$  is the major species. Its concentration is a maximum at the isoelectric point ( $\text{pI} = 5.97$ ). At pH = 9.60,  $[\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2^-] = [\text{H}_2\text{NCH}_2\text{CO}_2^-]$ . Above pH = 9.60,  $\text{H}_2\text{NCH}_2\text{CO}_2^-$  is the predominant species.

The equilibrium expressed by the preceding equation lies overwhelmingly to the side of the zwitterion.

Glycine, as well as other amino acids, is *amphoteric*, meaning it contains an acidic functional group and a basic functional group. The acidic functional group is the ammonium ion  $\text{H}_3\text{N}^+$ ; the basic functional group is the carboxylate ion  $-\text{CO}_2^-$ . How do we know this? Aside from its physical properties, the acid-base properties of glycine, as illustrated by the titration curve in Figure 25.1, require it. In a strongly acidic medium the species present is the cation  $\text{H}_3\text{N}^+\text{CH}_2\text{CO}_2\text{H}$ . As the pH is raised, its most acidic proton is removed. Is this proton removed from the positively charged nitrogen or from the carboxyl group? We know what to expect for the relative acid strengths of  $\text{RNH}_3^+$  and  $\text{RCO}_2\text{H}$ . A typical ammonium ion has  $\text{pK}_a \approx 9$ , and a typical carboxylic acid has  $\text{pK}_a \approx 5$ . The measured  $\text{pK}_a$  for the conjugate acid of glycine is 2.34, a value closer to that expected for deprotonation of the carboxyl group. As the pH is raised, a second deprotonation step, corresponding to removal of a proton from nitrogen of the zwitterion, is observed. The  $\text{pK}_a$  associated with this step is 9.60, much like that of typical alkylammonium ions.



Thus, glycine is characterized by two  $\text{pK}_a$  values: the one corresponding to the more acidic site is designated  $\text{pK}_{a1}$ , the one corresponding to the less acidic site is designated  $\text{pK}_{a2}$ . Table 25.2 lists  $\text{pK}_{a1}$  and  $\text{pK}_{a2}$  values for the  $\alpha$ -amino acids that have neutral side chains, which are the first two groups of amino acids given in Table 25.1. In all cases their  $\text{pK}_a$  values are similar to those of glycine.

Table 25.2 includes a column labeled  $\text{pI}$ , which is the *isoelectric point* of the amino acid. The **isoelectric point**, also called the **isoionic point**, is the pH at which the amino acid has no net charge. It is the pH at which the concentration of the zwitterion is a maximum. At a pH lower than  $\text{pI}$ , the amino acid is positively charged; at a pH higher than  $\text{pI}$ , the amino acid is negatively charged. For the amino acids in Table 25.2,  $\text{pI}$  is the average of  $\text{pK}_{a1}$  and  $\text{pK}_{a2}$  and lies slightly to the acid side of neutrality.

Some amino acids have side chains that bear acidic or basic groups. As Table 25.3 indicates, these amino acids are characterized by three  $\text{pK}_a$  values. The third  $\text{pK}_a$  reflects the nature of the side chain. Acidic amino acids (aspartic and glutamic acid) have acidic side chains; basic amino acids (lysine, arginine, and histidine) have basic side chains.

The isoelectric points of the amino acids in Table 25.3 are midway between the  $\text{pK}_a$  values of the zwitterion and its conjugate acid. Take two examples: aspartic acid and lysine.

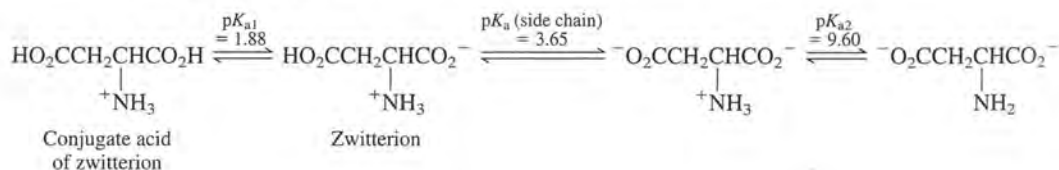
**TABLE 25.2** Acid–Base Properties of Amino Acids with Neutral Side Chains

| Amino acid    | $pK_{a1}^*$ | $pK_{a2}^*$ | pI   |
|---------------|-------------|-------------|------|
| Glycine       | 2.34        | 9.60        | 5.97 |
| Alanine       | 2.34        | 9.69        | 6.00 |
| Valine        | 2.32        | 9.62        | 5.96 |
| Leucine       | 2.36        | 9.60        | 5.98 |
| Isoleucine    | 2.36        | 9.60        | 6.02 |
| Methionine    | 2.28        | 9.21        | 5.74 |
| Proline       | 1.99        | 10.60       | 6.30 |
| Phenylalanine | 1.83        | 9.13        | 5.48 |
| Tryptophan    | 2.83        | 9.39        | 5.89 |
| Asparagine    | 2.02        | 8.80        | 5.41 |
| Glutamine     | 2.17        | 9.13        | 5.65 |
| Serine        | 2.21        | 9.15        | 5.68 |
| Threonine     | 2.09        | 9.10        | 5.60 |
| Tyrosine      | 2.20        | 9.11        | 5.66 |

\*In all cases,  $pK_{a1}$  corresponds to ionization of the carboxyl group;  $pK_{a2}$  corresponds to deprotonation of the ammonium ion.

Aspartic acid has an acidic side chain and a pI of 2.77. Lysine has a basic side chain and a pI of 9.74.

**Aspartic Acid:**



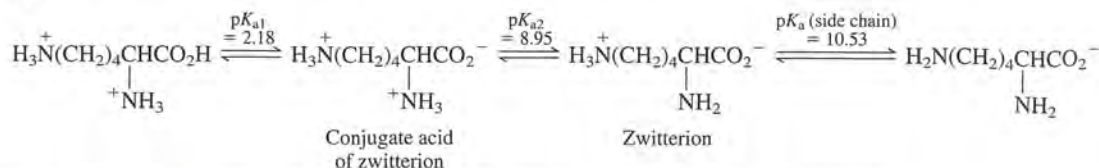
The pI of aspartic acid is the average of  $pK_{a1}$  (1.88) and the  $pK_a$  of the side chain (3.65) or 2.77.

**TABLE 25.3** Acid–Base Properties of Amino Acids with Ionizable Side Chains

| Amino acid    | $pK_{a1}^*$ | $pK_{a2}$ | $pK_a$ of side chain | pI    |
|---------------|-------------|-----------|----------------------|-------|
| Aspartic acid | 1.88        | 9.60      | 3.65                 | 2.77  |
| Glutamic acid | 2.19        | 9.67      | 4.25                 | 3.22  |
| Lysine        | 2.18        | 8.95      | 10.53                | 9.74  |
| Arginine      | 2.17        | 9.04      | 12.48                | 10.76 |
| Histidine     | 1.82        | 9.17      | 6.00                 | 7.59  |

\*In all cases,  $pK_{a1}$  corresponds to ionization of the carboxyl group of  $\text{RCHCO}_2\text{H}$  and  $pK_{a2}$  to ionization of the ammonium ion.

$\text{NH}_3^+$

**Lysine:**

The pI of lysine is the average of  $pK_{a2}$  (8.95) and the  $pK_a$  of the side chain (10.53) or 9.74.

**Problem 25.3**

Cysteine has  $pK_{a1} = 1.96$  and  $pK_{a2} = 10.28$ . The  $pK_a$  for ionization of the  $-\text{SH}$  group of the side chain is 8.18. What is the isoelectric point of cysteine?

**Problem 25.4**

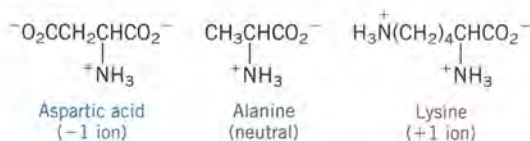
Above a pH of about 10, the major species present in a solution of tyrosine has a net charge of  $-2$ . Suggest a reasonable structure for this species.

Individual amino acids differ in their acid-base properties. This is important in peptides and proteins, where the properties of the substance depend on its amino acid constituents, especially on the nature of the side chains. It is also important in analyses in which a complex mixture of amino acids is separated into its components by taking advantage of the differences in their proton-donating and accepting power.

**Electrophoresis**

**E**lectrophoresis is a method for separation and purification that depends on the movement of charged particles in an electric field. Its principles can be introduced by considering some representative amino acids. The medium is a cellulose acetate strip that is moistened with an aqueous solution buffered at a particular pH. The opposite ends of the strip are placed in separate compartments containing the buffer, and each compartment is connected to a source of direct electric current (Figure 25.2a). If the buffer solution is more acidic than the isoelectric point (pI) of the amino acid, the amino acid has a net positive charge and migrates toward the negatively charged electrode. Conversely, when the buffer is more basic than the pI of the amino acid, the amino acid has a net negative charge and migrates toward the positively charged electrode. When the pH of the buffer corresponds to the pI, the amino acid has no net charge and does not migrate from the origin.

Thus if a mixture containing alanine, aspartic acid, and lysine is subjected to electrophoresis in a buffer that matches the isoelectric point of alanine (pH 6.0), aspartic acid (pI = 2.8) migrates toward the positive electrode, alanine remains at the origin, and lysine (pI = 9.7) migrates toward the negative electrode (Figure 25.2b).

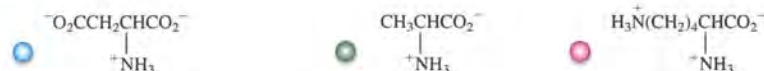


Electrophoresis is used primarily to analyze mixtures of peptides and proteins, rather than individual amino acids, but analogous principles apply. Because they incorporate different numbers of amino acids and because their side chains are different, two peptides will have slightly different acid-base properties and slightly different net charges at a particular pH. Thus, their mobilities in an electric field will be different, and electrophoresis can be used to separate them. The medium used to separate peptides and proteins is typically a polyacrylamide gel, leading to the term *gel electrophoresis* for this technique.

A second factor that governs the rate of migration during electrophoresis is the size (length and shape) of the peptide or protein. Larger molecules move through the polyacrylamide gel more slowly than smaller ones. In current practice, the experiment is modified to exploit differences in size more than differences in net charge, especially in the *SDS gel electrophoresis* of proteins. Approximately 1.5 g of the detergent *sodium dodecyl sulfate* (SDS, page 788) per gram of protein is added to the aqueous buffer. SDS binds to the protein, causing the protein to unfold so that it is roughly rod-shaped with the  $\text{CH}_3(\text{CH}_2)_{10}\text{CH}_2-$  groups of SDS associated with the lipophilic (hydrophobic) portions of the protein. The negatively charged sulfate groups are exposed to the water. The SDS molecules that they carry ensure that all the protein molecules are negatively charged and migrate toward the positive electrode. Further more, all the proteins in the mixture now have similar shapes and tend to travel at rates proportional

*Continued*

A mixture of amino acids



is placed at the center of a sheet of cellulose acetate. The sheet is soaked with an aqueous solution buffered at a pH of 6.0. At this pH aspartic acid (blue circle) exists as its  $-1$  ion, alanine (green circle) as its zwitterion, and lysine (pink circle) as its  $+1$  ion.



Application of an electric current causes the negatively charged ions to migrate to the  $+$  electrode, and the positively charged ions to migrate to the  $-$  electrode. The zwitterion, with a net charge of zero, remains at its original position.

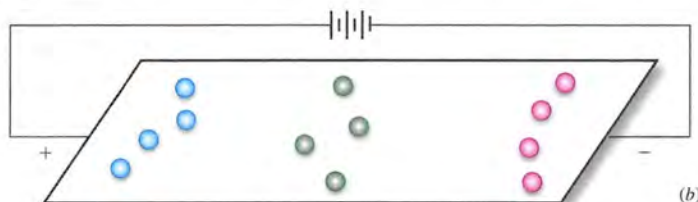


**Figure 25.2**

Application of electrophoresis to the separation of aspartic acid, alanine, and lysine according to their charge type at a pH corresponding to the isoelectric point (pI) of alanine.

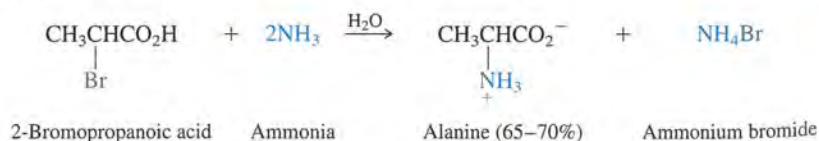
to their chain length. Thus, when carried out on a preparative scale, SDS gel electrophoresis permits proteins in a mixture to be separated according to their molecular weight. On an analytical scale, it is used to estimate the molecular weight of a protein

by comparing its electrophoretic mobility with that of proteins of known molecular weight.

Later, in Chapter 26, we will see how gel electrophoresis is used in nucleic acid chemistry.

## 25.4 Synthesis of Amino Acids

One of the earliest methods for the synthesis of amino acids dates back to the nineteenth century and is simply a nucleophilic substitution in which ammonia reacts with an  $\alpha$ -halo carboxylic acid.

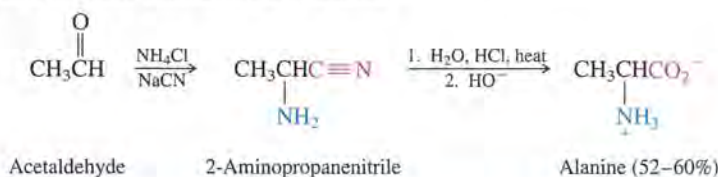


The  $\alpha$ -halo acid is normally prepared by the Hell-Volhard-Zelinsky reaction (Section 20.15).

### Problem 25.5

Outline the steps in a synthesis of valine from 3-methylbutanoic acid.

In the **Strecker synthesis** an aldehyde is converted to an  $\alpha$ -amino acid with one more carbon atom by a two-stage procedure in which an  $\alpha$ -amino nitrile is an intermediate. The  $\alpha$ -amino nitrile is formed by reaction of the aldehyde with ammonia or an ammonium salt and a source of cyanide ion. Hydrolysis of the nitrile group to a carboxylic acid function completes the synthesis.

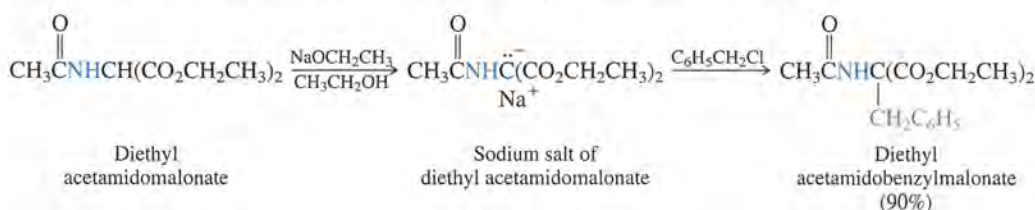


The synthesis of alanine was described by Adolf Strecker of the University of Würzburg (Germany) in a paper published in 1850.

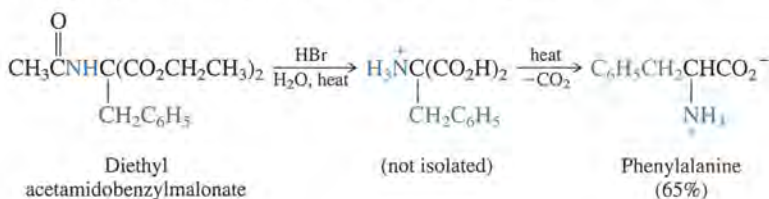
### Problem 25.6

Outline the steps in the preparation of valine by the Strecker synthesis.

A widely used method for the laboratory synthesis of  $\alpha$ -amino acids is a modification of the malonic ester synthesis (see Section 20.11). The key reagent is *diethyl acetamidomalonate*, a derivative of malonic ester that already has the critical nitrogen substituent in place at the  $\alpha$ -carbon atom. The side chain is introduced by alkylating diethyl acetamidomalonate in the same way as diethyl malonate itself is alkylated.



Hydrolysis removes the acetyl group from nitrogen and converts the two ester functions to carboxyl groups. Decarboxylation gives the desired product.



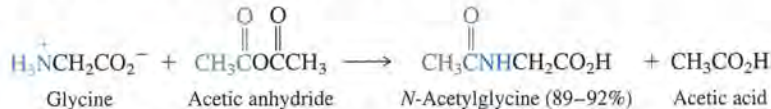
### Problem 25.7

Outline the steps in the synthesis of valine from diethyl acetamidomalonate. The overall yield of valine by this method is reported to be rather low (31%). Can you think of a reason why this synthesis is not very efficient?

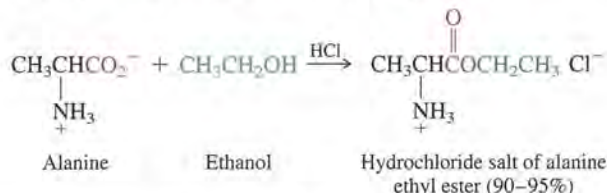
The  $\alpha$ -amino acids prepared by the synthetic methods just described are racemic unless a resolution step is included, enantiomerically enriched reactants are used, or the reaction is modified so as to become enantioselective. Considerable progress has been made in the last of these methods, allowing chemists to prepare not only L-amino acids, but also their much rarer D-enantiomers. We have already seen one example of this approach in the synthesis of the anti-parkinsonism drug L-dopa by enantioselective hydrogenation (see Section 14.14). A variation of the Strecker synthesis using a chiral catalyst has recently been developed that gives  $\alpha$ -amino acids with greater than 99% enantioselectivity.

## 25.5 Reactions of Amino Acids

Amino acids undergo reactions characteristic of both their amine and carboxylic acid functional groups. Acylation is a typical reaction of the amino group.

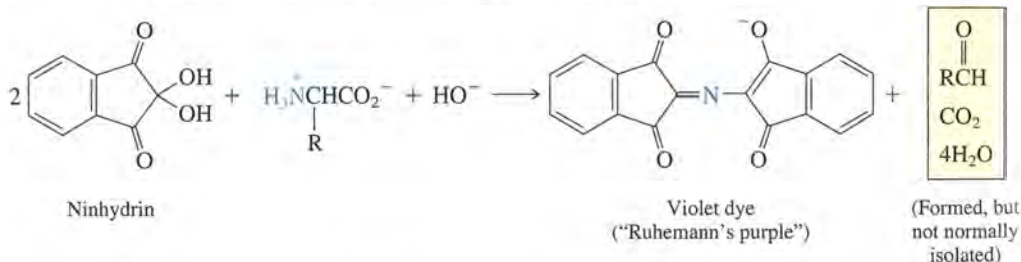


Ester formation is a typical reaction of the carboxyl group.



Ninhydrin is used to detect fingerprints.

The presence of amino acids can be detected by the formation of a purple color on treatment with *ninhydrin*. The same compound responsible for the purple color is formed from all amino acids in which the  $\alpha$ -amino group is primary.



Proline, in which the  $\alpha$ -amino group is secondary, gives an orange compound on reaction with ninhydrin.

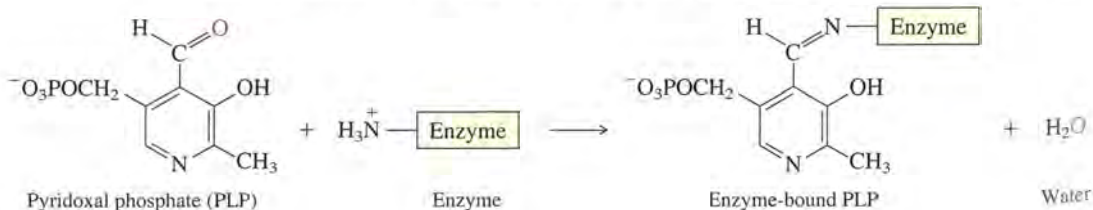
### Problem 25.8

Suggest a reasonable mechanism for the reaction of an  $\alpha$ -amino acid with ninhydrin.

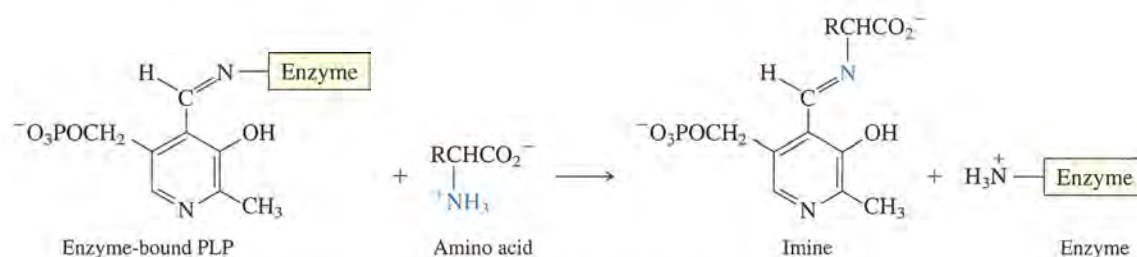
## 25.6 Some Biochemical Reactions of Amino Acids

The fact that we focus on the 20 (or 21 or 22) amino acids that are coded for by DNA indicates the importance we place on protein biosynthesis as a reaction of amino acids. In addition to serving as building blocks for proteins, amino acids are involved in numerous other biochemical processes. They store energy, although less efficiently than carbohydrates and lipids, and are starting materials for the biosynthesis of other amino acids, amines, alkaloids, and neurotransmitters.

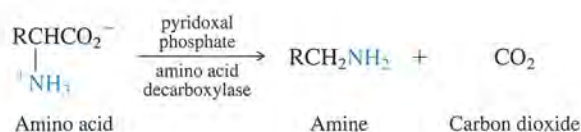
Many of the biochemical reactions of amino acids require *pyridoxal 5'-phosphate* (PLP), the active form of vitamin B<sub>6</sub>, as a coenzyme. Before acting on an amino acid, PLP uses its aldehyde function to form an imine with the amino group of a lysine side chain of a protein.



Reaction of the enzyme-bound PLP with an amino acid connects the amino acid and PLP by a new imine linkage.



The pyridine ring of PLP, especially when protonated, facilitates several kinds of reactions at the amino acid's  $\alpha$  carbon by acting as an electron-withdrawing group. One is decarboxylation.

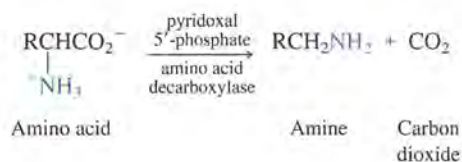


Mechanism 25.1 outlines the mechanism of decarboxylation, showing the role played by the coenzyme.

### Mechanism 25.1

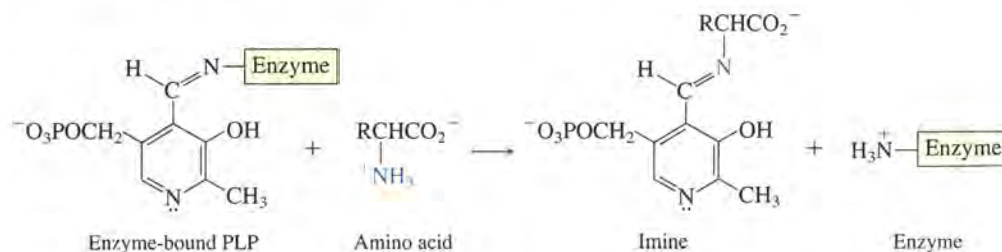
#### Pyridoxal 5'-Phosphate-Mediated Decarboxylation of an $\alpha$ -Amino Acid

##### THE OVERALL REACTION:



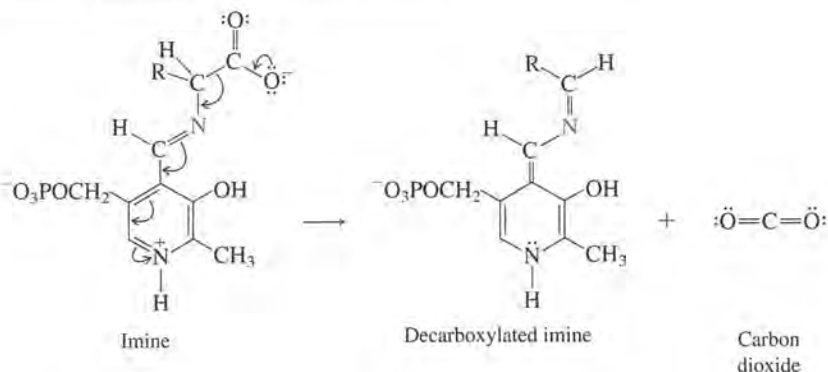
**THE MECHANISM:** Each stage is enzyme-catalyzed and can involve more than one elementary step.

**Stage 1:** The amino acid reacts with enzyme-bound pyridoxal 5'-phosphate (PLP). An imine linkage (C=N) between the amino acid and PLP forms, and the enzyme is displaced.

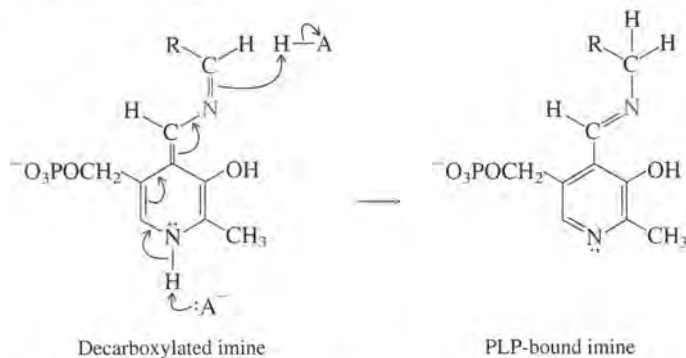


*Continued*

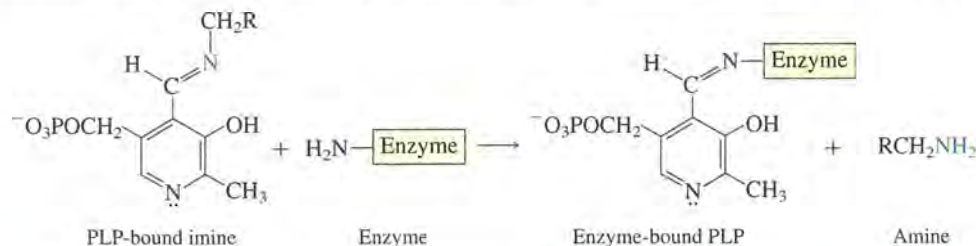
**Stage 2:** When the pyridine ring is protonated on nitrogen, it becomes a stronger electron-withdrawing group, and decarboxylation is facilitated by charge neutralization.



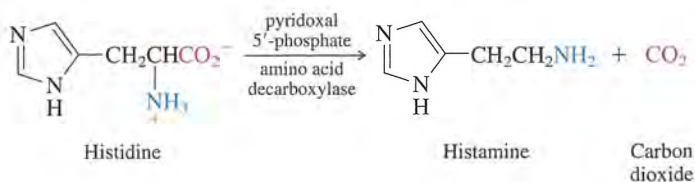
**Stage 3:** Proton transfer to the  $\alpha$  carbon and abstraction of a proton from the pyridine nitrogen brings about rearomatization of the pyridine ring.



**Stage 4:** Reaction of the PLP-bound imine with the enzyme liberates the amine and restores the enzyme-bound coenzyme.



Many bioactive amines arise by PLP-assisted amino acid decarboxylation. Decarboxylation of histidine, for example, gives histamine, a powerful vasodilator normally present in the body but formed in excessive amounts under conditions of traumatic shock.



Histamine is present in various tissues and produces different effects depending on the kind of receptor it binds to. Binding of histamine to  $H_1$  receptors in mast cells triggers, for example, the sneezing and watery eyes of hay fever and the itching of mosquito bites. The  $H_2$  receptors in the cells that line the stomach regulate the secretion of gastric acid. The present generation of antiallergy and antiulcer drugs bind to  $H_1$  and  $H_2$ , respectively, and act by denying histamine access to these receptors.

### Problem 25.9

One of the amino acids in Table 25.1 is the biological precursor to  $\gamma$ -aminobutyric acid (4-aminobutanolic acid), which it forms by a decarboxylation reaction. Which amino acid is this?

The chemistry of the brain and central nervous system is affected by a group of substances called **neurotransmitters**, substances that carry messages across a synapse from one neuron to another. Several of these neurotransmitters arise from L-tyrosine by structural modification and decarboxylation, as outlined in Figure 25.3.

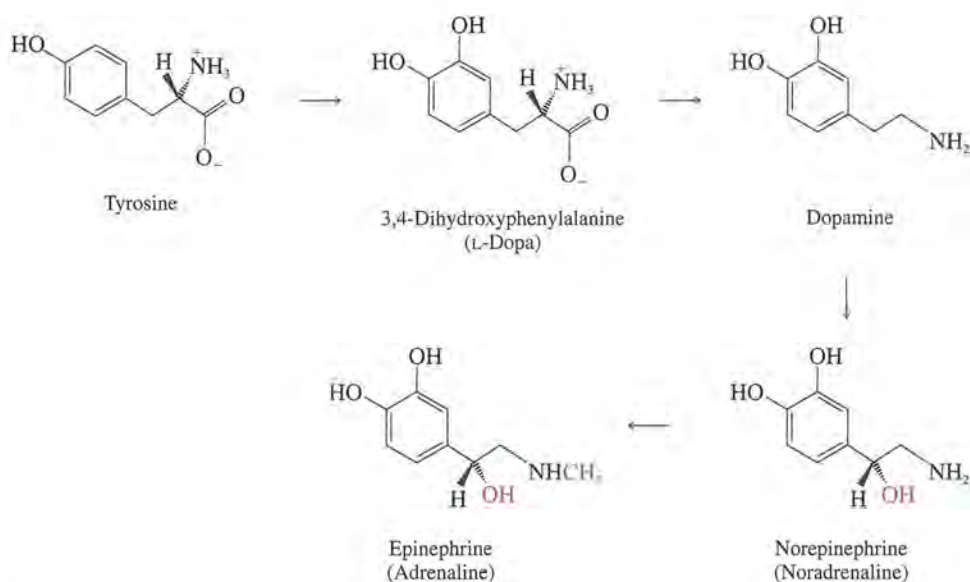


Figure 25.3

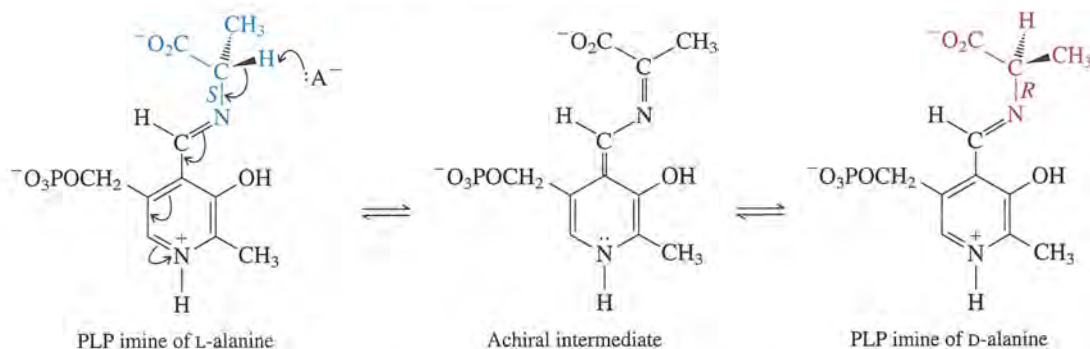
Tyrosine is the biosynthetic precursor to a number of neurotransmitters. Each transformation is enzyme-catalyzed.

### Problem 25.10

Which of the transformations in Figure 25.3 is catalyzed by an amino acid decarboxylase?

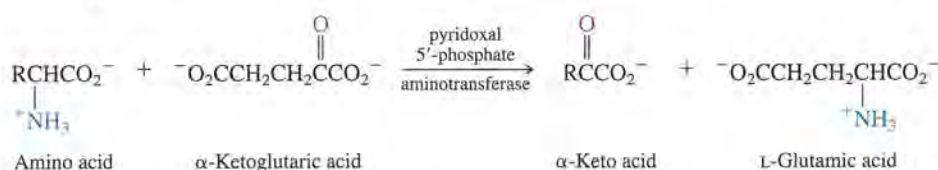
Many of the drugs prescribed to treat anxiety, depression, or attention deficit disorder are “reuptake” inhibitors. They increase the concentration in the brain of a necessary neurotransmitter such as dopamine or epinephrine by slowing the rate at which it is reabsorbed.

Pyridoxal 5'-phosphate is also a coenzyme for the enzyme-catalyzed racemization of amino acids. The key reaction is proton abstraction from the  $\alpha$  carbon of the amino acid imine of PLP. This step converts the  $\alpha$  carbon, which is a chirality center, from  $sp^3$  to  $sp^2$ .

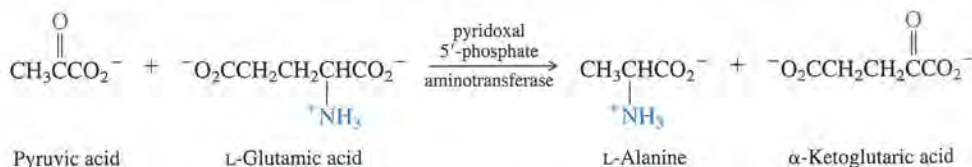


Proton transfer to the imine carbon of the achiral intermediate gives equal amounts of both enantiomers of the PLP imine. The equation illustrates the racemization of L-alanine, which is catalyzed by the PLP-dependent enzyme *alanine racemase*. Because D-alanine is an essential component of bacterial cell walls, there is considerable interest in designing inhibitors of alanine racemase as potential antibacterial drugs.

In addition to amino acid decarboxylation and racemization, PLP is a coenzyme for **transamination**—the transfer of an amino group from one compound to another. The enzymes that catalyze transaminations are called *aminotransferases* or *transaminases*. Many transaminations involve two compounds:  $\alpha$ -ketoglutaric acid and L-glutamic acid.



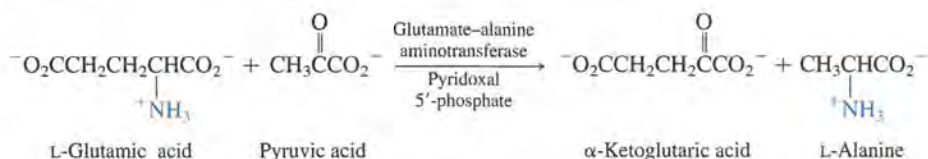
The reaction shown, written in the forward direction illustrates a feature of amino acid metabolism, the breaking down of amino acids and using their structural units for other purposes. Written in the reverse direction, it illustrates a biosynthetic pathway to amino acids. L-Alanine, for example, is not an essential amino acid because we have the capacity to biosynthesize it. One biosynthetic route to L-alanine is the transamination of pyruvic acid.



Although the equation shows a single transamination, the mechanism actually involves two, each of which is understandable in light of what we have already seen concerning the way PLP functions as a coenzyme. In the first, as outlined in the first four stages of Mechanism 25.2, the amino group of L-glutamic acid is transferred to the coenzyme PLP to give pyridoxamine 5'-phosphate (PMP). The second transamination shown in abbreviated form as stage 5, is analogous to the first but followed in reverse order.

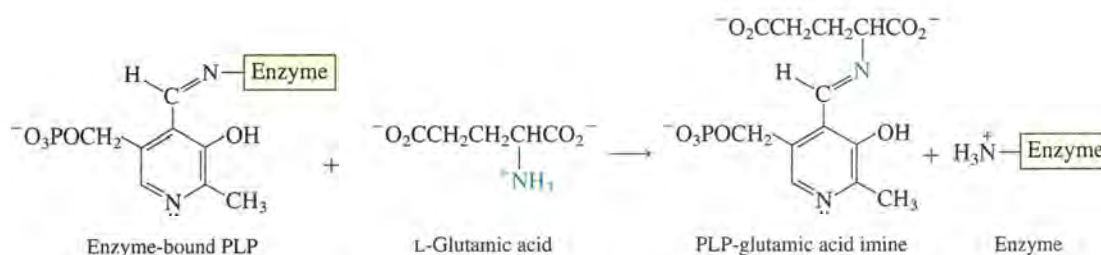
### Problem 25.11

$\alpha$ -Ketoglutaric acid undergoes a transamination reaction with L-aspartic acid (see Table 25.1), converting it to a compound known as oxaloacetic acid. What is the structure of oxaloacetic acid?

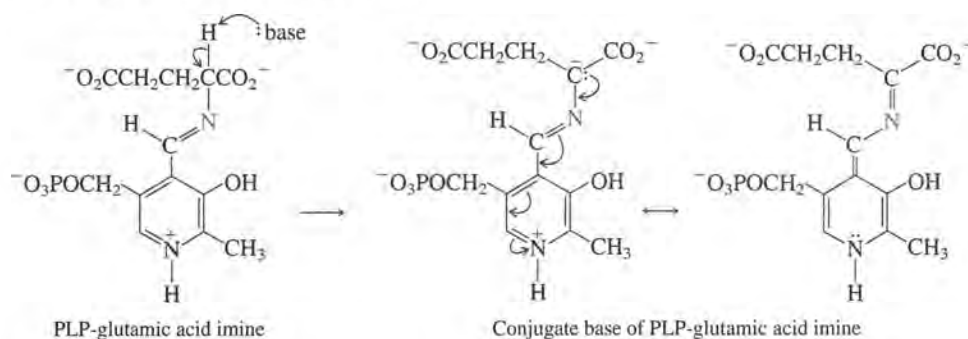
**Mechanism 25.2****Transamination: Biosynthesis of L-Alanine from L-Glutamic Acid and Pyruvic Acid****THE OVERALL REACTION:**

**THE MECHANISM:** Each stage can involve more than one elementary step. Each reaction is enzyme-catalyzed. Stages 1–4 show the transfer of the amino group of L-glutamic acid to pyridoxal 5'-phosphate to give  $\alpha$ -ketoglutaric acid and pyridoxamine 5'-phosphate (PMP). PMP reacts with pyruvic acid to give an imine, which then follows stages analogous to 1–4, but in reverse order, to give L-alanine and PLP. These stages are summarized as stage 5.

**Stage 1:** L-Glutamic acid forms an imine bond to the coenzyme PLP by reaction with imine formed between PLP and the enzyme.



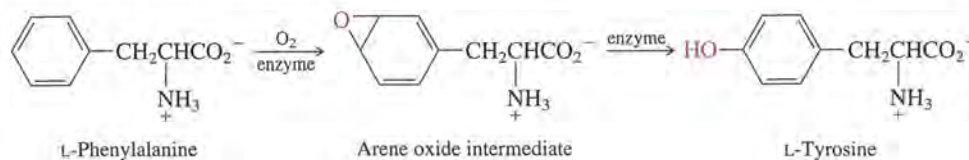
**Stage 2:** The electron-withdrawing effect of the pyridinium ring stabilizes the conjugate base formed by proton abstraction from the  $\alpha$  carbon of the imine.



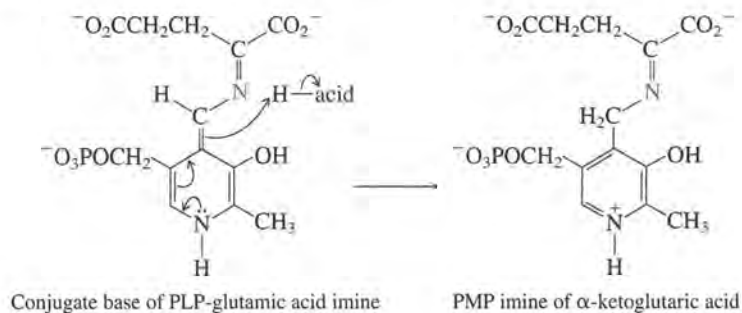
Continued

Peptide-bond formation and transamination are the most general reactions of the standard amino acids, but individual amino acids often undergo reactions of more limited scope. One of the biosynthetic pathways to L-tyrosine is oxidation of L-phenylalanine. An *arene oxide* is an intermediate.

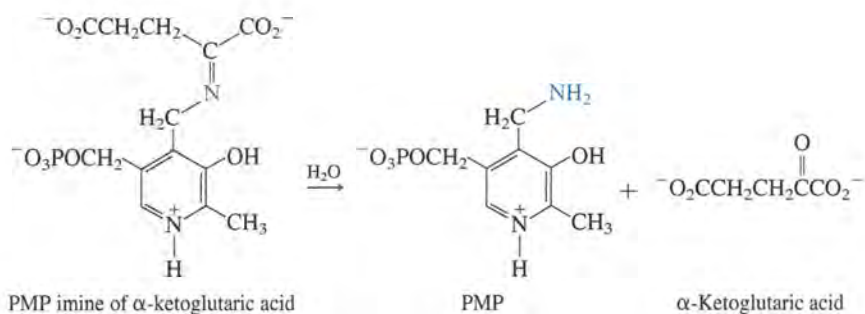
For more on this reaction, see Descriptive Passage and Interpretive Problems 16: Epoxide Rearrangements and the NIH Shift.



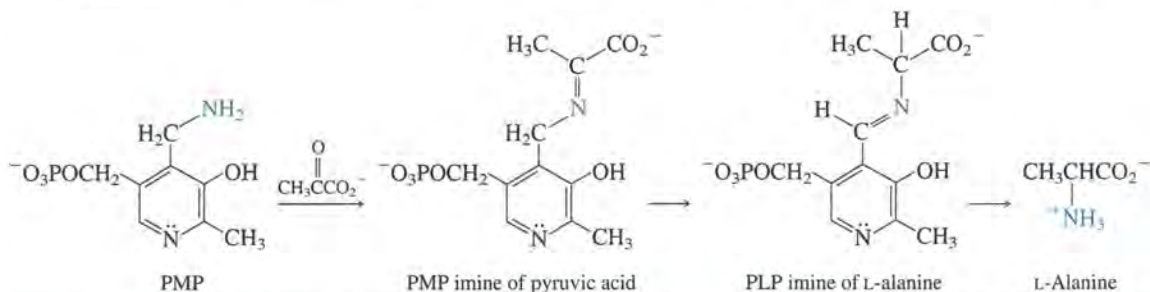
**Stage 3:** Electron reorganization and protonation of carbon restores the aromaticity of the pyridine ring while converting a PLP imine to a PMP imine.



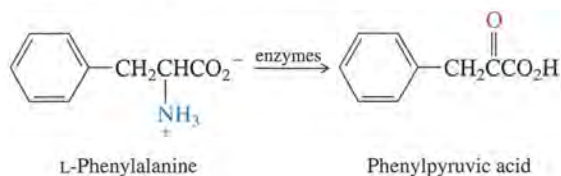
**Stage 4:** Cleavage of the PMP imine, shown here as a hydrolysis, gives pyridoxamine and  $\alpha$ -ketoglutaric acid.



**Stage 5:** Formation of the imine from PMP and pyruvic acid sets the stage for the conversion of pyruvic acid to L-alanine.



Some individuals lack the enzyme *phenylalanine hydroxylase* required for this conversion, and any L-phenylalanine that would ordinarily be converted to L-tyrosine is converted to phenylpyruvic acid by transamination.

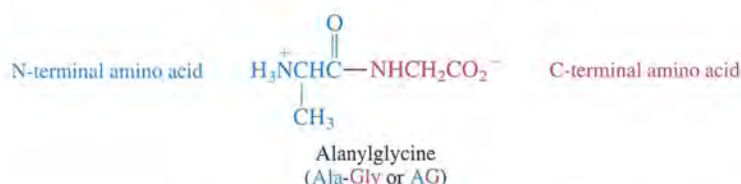


Too much phenylpyruvic acid causes *phenylketonuria* (PKU disease), which can lead to mental retardation in growing children. Infants are routinely screened for PKU disease within a few days of birth. PKU disease cannot be cured, but is controlled by restricting the dietary intake of foods, such as meat, that are rich in L-phenylalanine.

Foods sweetened with aspartame (page 1048) contain a PKU warning. Can you see why?

## 25.7 Peptides

A key biochemical reaction of amino acids is their conversion to peptides, polypeptides, and proteins. In all these substances amino acids are linked together by amide bonds. The amide bond between the amino group of one amino acid and the carboxyl of another is called a **peptide bond**. Alanine-glycine is a representative dipeptide.



It is understood that  $\alpha$ -amino acids occur as their L stereoisomers unless otherwise indicated. The D notation is explicitly shown when a D amino acid is present, and a racemic amino acid is identified by the prefix DL.

By agreement, peptide structures are written so that the amino group (as  $\text{H}_3\text{N}^+$ — or  $\text{H}_2\text{N}$ —) is at the left and the carboxyl group (as  $\text{CO}_2^-$  or  $\text{CO}_2\text{H}$ ) is at the right. The left and right ends of the peptide are referred to as the **N terminus** (or amino terminus) and the **C terminus** (or carboxyl terminus), respectively. Alanine is the N-terminal amino acid in alanylglycine; glycine is the C-terminal amino acid. A dipeptide is named as an acyl derivative of the C-terminal amino acid. The precise order of bonding in a peptide (its amino acid *sequence*) is conveniently specified by using the three-letter amino acid abbreviations for the respective amino acids and connecting them by hyphens. One-letter abbreviations are used without punctuation. Individual amino acid components of peptides are often referred to as **amino acid residues**.

### Problem 25.12

Write structural formulas showing the constitution of each of the following dipeptides. Rewrite each sequence using one-letter abbreviations for the amino acids.

- |             |                 |
|-------------|-----------------|
| (a) Gly-Ala | (d) Gly-Glu     |
| (b) Ala-Phe | (e) Lys-Gly     |
| (c) Phe-Ala | (f) D-Ala-D-Ala |

**Sample Solution** (a) Gly-Ala is a constitutional isomer of Ala-Gly. Glycine is the N-terminal amino acid in Gly-Ala; alanine is the C-terminal amino acid.

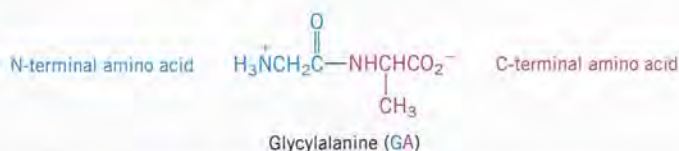
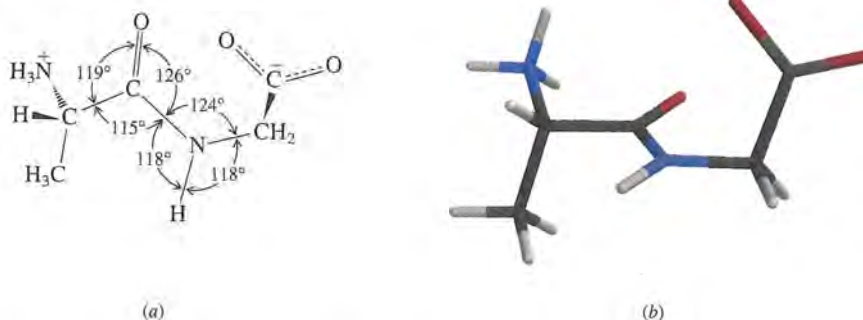


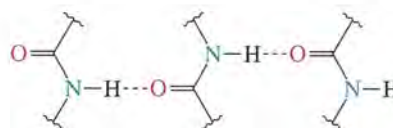
Figure 25.4 shows the structure of Ala-Gly as determined by X-ray crystallography. An important feature is the planar geometry of the peptide bond, and the most stable conformation with respect to this bond has the two  $\alpha$ -carbon atoms anti to each other. Rotation about the amide bond is slow because delocalization of the unshared electron pair of nitrogen into the carbonyl group gives partial double-bond character to the carbon–nitrogen bond.

Figure 25.4

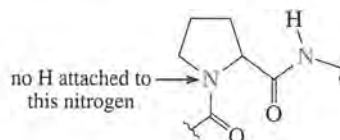
Structural features of the dipeptide L-alanylglycine as determined by X-ray crystallography. All of the bonds of the peptide linkage lie in the same plane and both  $\alpha$  carbons are anti to each other.



In addition to its planar geometry, the amide bond affects the structure of peptides in another important way. The N—H and the C=O units are candidates for hydrogen bonding with other peptide linkages both within the same and with adjacent polypeptide chains.



As the only secondary amine among the standard amino acids, L-proline is an exception in that its amides lack an N—H bond.

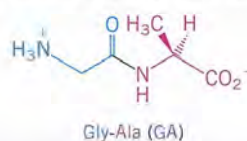


This structural feature of L-proline affects the three-dimensional shape of peptides that contain it by limiting the number of hydrogen-bonding opportunities.

### Problem 25.13

Expand your answer to Problem 25.12 by showing the structural formula for each dipeptide in a manner that reveals the stereochemistry at the  $\alpha$ -carbon atom.

**Sample Solution** (a) Glycine is achiral, and so Gly-Ala has only one chirality center, the  $\alpha$ -carbon atom of the L-alanine residue. When the carbon chain is drawn in an extended zigzag fashion and L-alanine is the C terminus, its structure is as shown:



The structures of higher peptides are extensions of the structural features of dipeptides. Figure 25.5 gives the structural formula and amino acid sequence of the naturally occurring pentapeptide *leucine enkephalin*. Enkephalins are pentapeptide components of *endorphins*, polypeptides present in the brain that act as the body's own painkillers. A second substance, *methionine enkephalin*, is also present in endorphins. Methionine enkephalin is about 20 times more potent than leucine enkephalin. It differs from leucine enkephalin only in having methionine instead of leucine as its C-terminal amino acid.

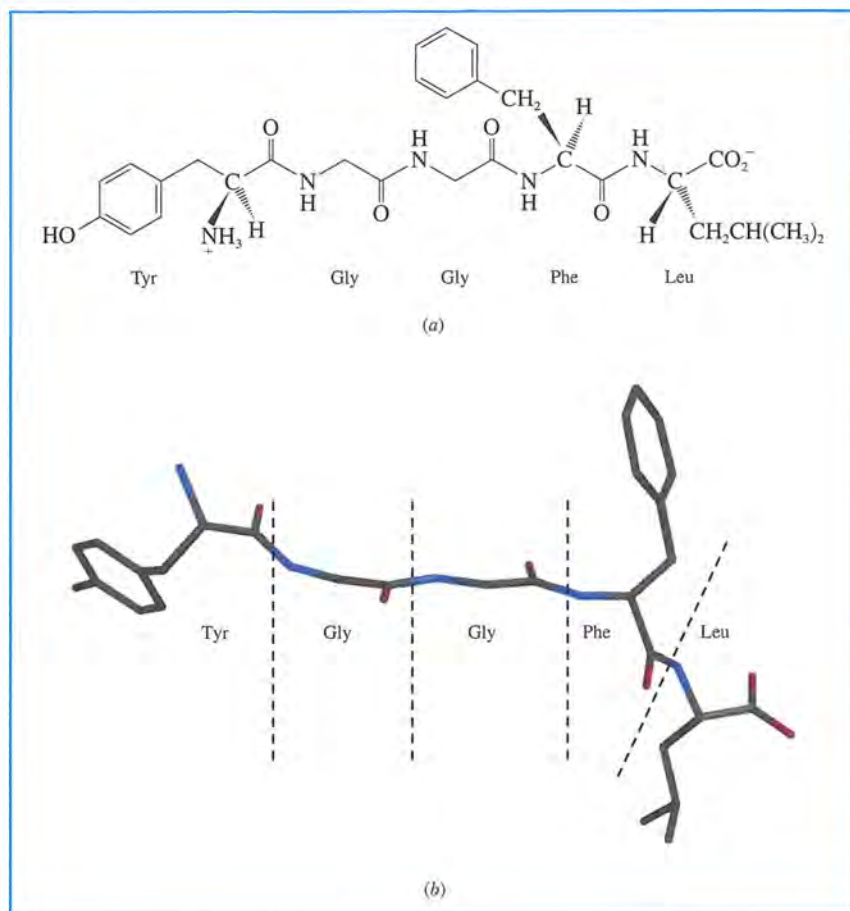


Figure 25.5

The structure of the pentapeptide leucine enkephalin shown as (a) a structural drawing and (b) as a molecular model. The shape of the molecular model was determined by X-ray crystallography. Hydrogens have been omitted for clarity.

## Problem 25.14

What is the amino acid sequence (using three-letter abbreviations) of methionine enkephalin? Also show it using one-letter abbreviations.

Peptides having structures slightly different from those described to this point are known. One such variation is seen in the nonapeptide *oxytocin*, shown in Figure 25.6. Oxytocin is a hormone secreted by the pituitary gland that stimulates uterine contractions during childbirth and promotes lactation. Rather than terminating in a carboxyl group, the C-terminal glycine residue in oxytocin has been modified to become its corresponding amide. Two

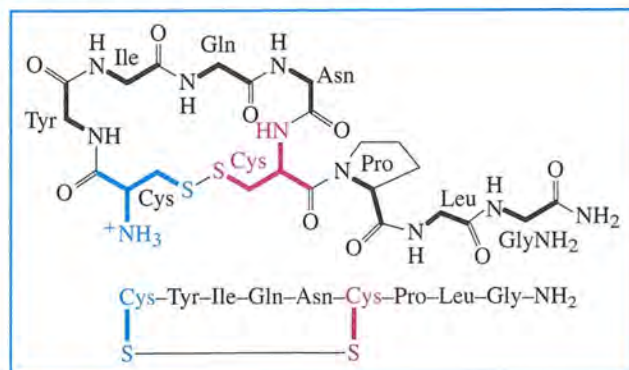


Figure 25.6

The connectivity of oxytocin with most of the side chains omitted for clarity. A disulfide bond connects the two cysteines. The cysteine shown in blue is the N-terminal amino acid; the one in red is the fourth amino acid beginning at the C-terminus. The C-terminus is the amide of glycine.

Recall from Section 15.12 that compounds of the type RSH are readily oxidized to RSSR.

cysteine units, one of them the N-terminal amino acid, are joined by the sulfur–sulfur bond of a large-ring cyclic disulfide unit. This is a common structural modification in polypeptides and proteins that contain cysteine residues. It provides a covalent bond between regions of peptide chains that may be many amino acid residues removed from each other.

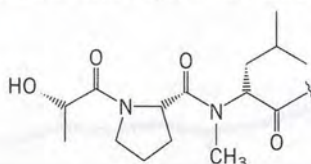
### Problem 25.15

What is the net charge of oxytocin at pH = 7?

Cyclic peptides in which the backbone is made entirely of amide bonds are also well known. Examples include the didemnins, which were featured in our chapter opener.

### Problem 25.16

The side chain in didemnin B has two amino acids that resemble those in Table 25.1. Which two? How are they different from those in the table?



Sanger was a corecipient of a second Nobel Prize in 1980 for devising methods for sequencing nucleic acids. Sanger's strategy for nucleic acid sequencing will be described in Section 26.14.

## 25.8 Introduction to Peptide Structure Determination

There are several levels of peptide structure. The **primary structure** is the amino acid sequence plus any disulfide links. With the 20 amino acids of Table 25.1 as building blocks,  $20^2$  dipeptides,  $20^3$  tripeptides,  $20^4$  tetrapeptides, and so on, are possible. Given a peptide of unknown structure, how do we determine its amino acid sequence?

We'll describe peptide structure determination by first looking at one of the great achievements of biochemistry, the determination of the amino acid sequence of insulin by Frederick Sanger of Cambridge University (England). Sanger was awarded the 1958 Nobel Prize in Chemistry for this work, which he began in 1944 and completed 10 years later. The methods used by Sanger and his coworkers are, of course, dated by now, but the overall logic hasn't changed very much. We'll use Sanger's insulin strategy to orient us, then show how current methods of protein sequencing have evolved from it.

Sanger's strategy can be outlined as follows:

1. Determine what amino acids are present and their molar ratios.
2. Cleave the peptide into smaller fragments, separate these fragments, and determine the amino acid composition of the fragments.
3. Identify the N-terminal and the C-terminal amino acid in the original peptide and in each fragment.
4. Organize the information so that the amino acid sequences of small fragments can be overlapped to reveal the full sequence.

## 25.9 Amino Acid Analysis

The chemistry behind amino acid analysis begins with acid-catalyzed hydrolysis of amide bonds. The peptide is hydrolyzed by heating in 6 M hydrochloric acid for about 24 h to give a solution that contains all of the amino acids. Analysis of the mixture in terms of its components and their relative amounts is typically done by chromatographic methods.

These methods flow from the work of Stanford Moore and William H. Stein of Rockefeller University who developed automated techniques for separating and identifying amino acids. In their original work, Moore and Stein used ion-exchange chromatography. Modern methods based on high-performance liquid chromatography (HPLC) are both faster and more selective for separating the individual amino acids in a mixture. Either before or after their separation, the amino acids are allowed to react ("tagged") with a substance that bears a group—a naphthalene ring, for example that fluoresces. The fluorescence is strong enough so that modern analyzers can detect the amino acids obtained from  $10^{-5}$  to  $10^{-7}$ g of peptide.

Moore and Stein shared one half of the 1972 Nobel Prize in Chemistry.

Fluorescence is the emission of radiation by a substance after it has absorbed radiation of a higher frequency.

### Problem 25.17

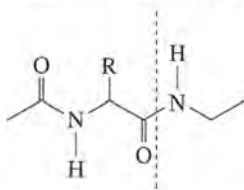
Amino acid analysis of a certain tetrapeptide gave alanine, glycine, phenylalanine, and valine in equimolar amounts. What amino acid sequences are possible for this tetrapeptide?

## 25.10 Partial Hydrolysis of Peptides

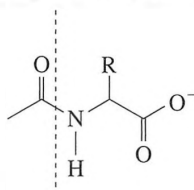
Whereas acid-catalyzed hydrolysis of peptides cleaves amide bonds indiscriminately and eventually breaks all of them, enzymatic hydrolysis is much more selective and is the method used to convert a peptide into smaller fragments.

The enzymes that catalyze the hydrolysis of peptide bonds are called **peptidases**. *Trypsin*, a digestive enzyme present in the intestine, catalyzes only the hydrolysis of peptide bonds involving the carboxyl group of a lysine or arginine residue. *Chymotrypsin*, another digestive enzyme, is selective for peptide bonds involving the carboxyl group of amino acids with aromatic side chains (phenylalanine, tyrosine, tryptophan). One group of pancreatic enzymes, known as *carboxypeptidases*, catalyzes only the hydrolysis of the peptide bond to the C-terminal amino acid. In addition to these, many other digestive enzymes are known and their selectivity exploited in the selective hydrolysis of peptides.

Papain, the active component of most meat tenderizers, is a peptidase.



Trypsin cleaves here when  
R = side chain of lysine or arginine



Carboxypeptidase cleaves the peptide  
bond of the C-terminal amino acid

Chymotrypsin cleaves here when  
R = side chain of phenylalanine, tyrosine, or tryptophan

### Problem 25.18

Digestion of the tetrapeptide of Problem 25.17 with chymotrypsin gave a dipeptide that on amino acid analysis gave phenylalanine and valine in equimolar amounts. What amino acid sequences are possible for the tetrapeptide?

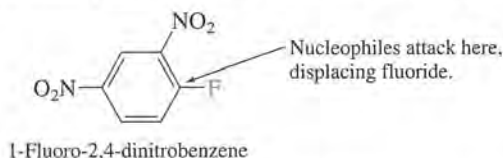
## 25.11 End Group Analysis

An amino acid sequence is ambiguous unless we know the direction in which to read it—left to right, or right to left. We need to know which end is the N terminus and which is the C terminus. As we saw in the preceding section, carboxypeptidase-catalyzed

hydrolysis cleaves the C-terminal amino acid and so can be used to identify it. What about the N terminus?

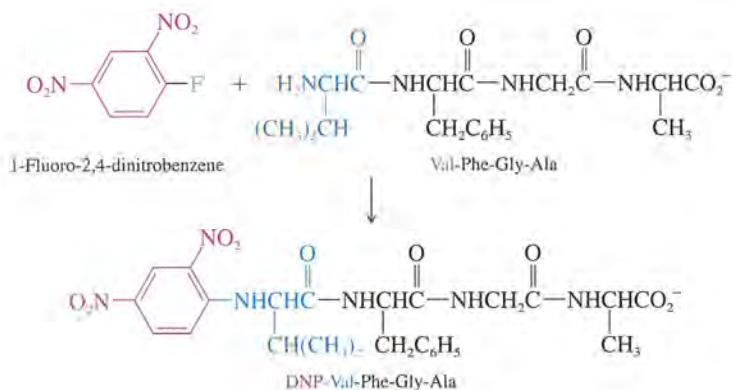
Several chemical methods have been devised for identifying the N-terminal amino acid. They all take advantage of the fact that the N-terminal amino group is free and can act as a nucleophile. The  $\alpha$ -amino groups of all the other amino acids are part of amide linkages, are not free, and are much less nucleophilic. Sanger's method for N-terminal residue analysis involves treating a peptide with 1-fluoro-2,4-dinitrobenzene, which is very reactive toward nucleophilic aromatic substitution (see Chapter 12).

1-Fluoro-2,4-dinitrobenzene is commonly referred to as *Sanger's reagent*.



The amino group of the N-terminal amino acid displaces fluoride from 1-fluoro-2,4-dinitrobenzene and gives a peptide in which the N-terminal nitrogen is labeled with a 2,4-dinitrophenyl (DNP) group. This is shown for the case of Val-Phe-Gly-Ala in Figure 25.7. The 2,4-dinitrophenyl-labeled peptide DNP-Val-Phe-Gly-Ala is isolated and subjected to hydrolysis, after which the 2,4-dinitrophenyl derivative of the N-terminal amino acid is isolated and identified as DNP-Val by comparing its chromatographic behavior with that of standard samples of 2,4-dinitrophenyl-labeled amino acids. None

The reaction is carried out by mixing the peptide and 1-fluoro-2,4-dinitrobenzene in the presence of a weak base such as sodium carbonate. In the first step the base abstracts a proton from the terminal  $H_3N$  group to give a free amino function. The nucleophilic amino group attacks 1-fluoro-2,4-dinitrobenzene, displacing fluoride.



Acid hydrolysis cleaves the amide bonds of the 2,4-dinitrophenyl-labeled peptide, giving the 2,4-dinitrophenyl-labeled N-terminal amino acid and a mixture of unlabeled amino acids.

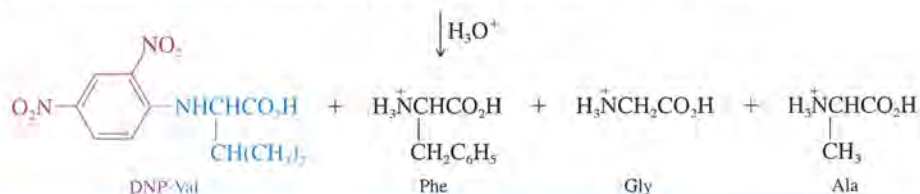


Figure 25.7

Use of 1-fluoro-2,4-dinitrobenzene to identify the N-terminal amino acid of a peptide.

of the other amino acid residues bear a 2,4-dinitrophenyl group; they appear in the hydrolysis product as the free amino acids.

Labeling the N-terminal amino acid as its DNP derivative is mainly of historical interest and has been replaced by other methods. We'll discuss one of these—the Edman degradation—in Section 25.13. First, though, we'll complete our review of the general strategy for peptide sequencing by seeing how Sanger tied all of the information together into a structure for insulin.

## 25.12 Insulin

Sanger worked with insulin from cows, which has 51 amino acids, divided between two chains. One of these, the A chain, has 21 amino acids; the other, the B chain, has 30. The A and B chains are joined by disulfide bonds between cysteine residues (Cys-Cys). Figure 25.8 shows some of the information that defines the amino acid sequence of the B chain.

- Reaction of the B chain peptide with 1-fluoro-2,4-dinitrobenzene established that phenylalanine is the N terminus.
- Pepsin-catalyzed hydrolysis gave the four peptides shown in blue in Figure 25.8. (Their sequences were determined in separate experiments.) These four peptides contain 27 of the 30 amino acids in the B chain, but there are no points of overlap between them.
- The sequences of the four tetrapeptides shown in red in Figure 25.8 bridge the gaps between three of the four “blue” peptides to give an unbroken sequence from 1 through 24.
- The peptide shown in green was isolated by trypsin-catalyzed hydrolysis and has an amino acid sequence that completes the remaining overlaps.

The collection of sequenced fragments constitutes the **peptide map** for insulin.

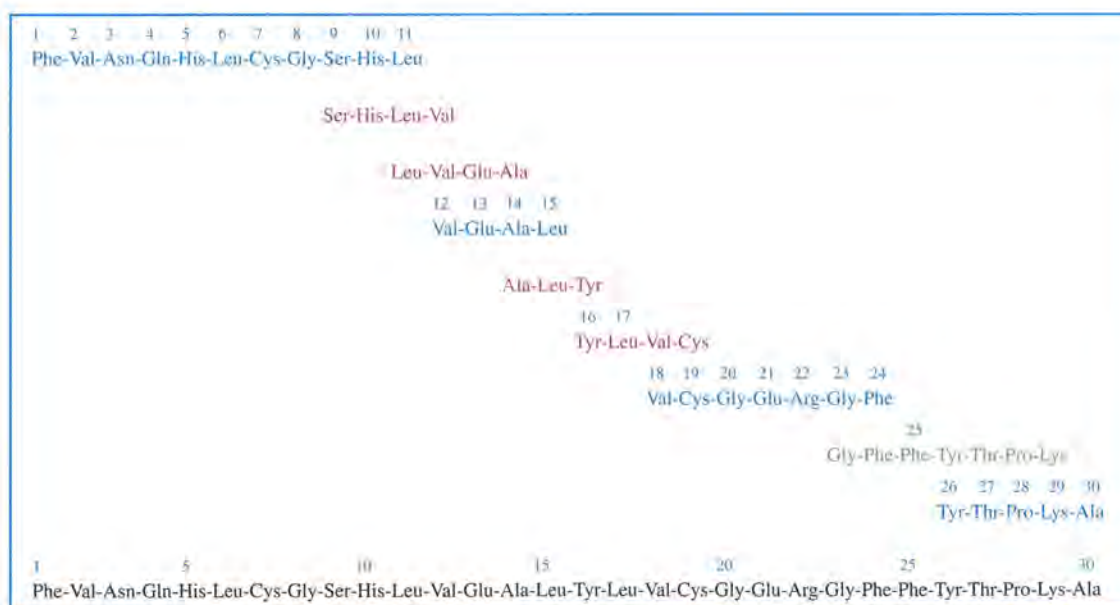
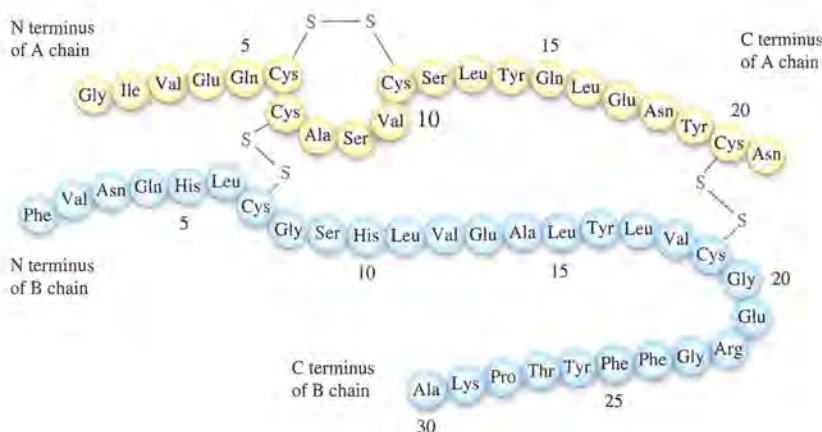


**Figure 25.8**

Diagram showing how the amino acid sequence of the B chain of bovine insulin can be determined by overlap of peptide fragments. Pepsin-catalyzed hydrolysis produced the fragments shown in blue, trypsin produced the one shown in green, and acid-catalyzed hydrolysis gave many fragments, including the four shown in red.

Figure 25.9

The amino acid sequence in bovine insulin. The A chain is joined to the B chain by two disulfide units (shown in green). There is also a disulfide bond linking cysteines 6 and 11 in the A chain. Human insulin has threonine and isoleucine at residues 8 and 10, respectively, in the A' chain and threonine as the C-terminal amino acid in the B chain.



Sanger also determined the sequence of the A chain and identified the cysteine residues involved in disulfide bonds between the A and B chains as well as in the disulfide linkage within the A chain. The complete insulin structure is shown in Figure 25.9. The structure shown is that of bovine insulin. The A chains of human insulin and bovine insulin differ in only two amino acid residues; their B chains are identical except for the amino acid at the C terminus.

### 25.13 The Edman Degradation and Automated Sequencing of Peptides

When Sanger's method for N-terminal residue analysis was discussed, you may have wondered why it was not done sequentially. Simply start at the N terminus and work steadily back to the C terminus identifying one amino acid after another. The idea is fine, but it just doesn't work well in practice, at least with 1-fluoro-2,4-dinitrobenzene.

A major advance was devised by Pehr Edman (University of Lund, Sweden) that has become the standard method for N-terminal residue analysis. The **Edman degradation** is based on the chemistry shown in Mechanism 25.3. A peptide reacts with phenyl isothiocyanate to give a *phenylthiocarbamoyl* (PTC) derivative, as shown in the first step. This PTC derivative is then treated with an acid in an *anhydrous* medium (Edman used nitromethane saturated with hydrogen chloride) to cleave the amide bond between the N-terminal amino acid and the remainder of the peptide. No other peptide bonds are cleaved in this step as amide bond hydrolysis requires water. When the PTC derivative is treated with acid in an anhydrous medium, the sulfur atom of the C=S unit acts as an internal nucleophile, and the only amide bond cleaved under these conditions is the one to the N-terminal amino acid. The product of this cleavage, called a *thiazolone*, is unstable under the conditions of its formation and rearranges to a *phenylthiohydantoin* (PTH), which is isolated and identified by comparing it with standard samples of PTH derivatives of known amino acids. This is normally done by chromatographic methods, but mass spectrometry has also been used.

Only the N-terminal amide bond is broken in the Edman degradation; the rest of the peptide chain remains intact. It can be isolated and subjected to a second Edman procedure to determine its new N terminus. We can proceed along a peptide chain by beginning with the N terminus and determining each amino acid in order. The sequence is given directly by the structure of the PTH derivative formed in each successive degradation.

#### Problem 25.19

Give the structure of the PTH derivative isolated in the second Edman cycle of the tetrapeptide Val-Phe-Gly-Ala.



## Peptide Mapping and MALDI Mass Spectrometry

**B**iological materials often contain proteins that must be identified. Recent advances in mass spectrometry have made peptide mapping a convenient tool for this purpose. The protein in question is selectively hydrolyzed with a peptidase such as trypsin and the mixture of peptides produced is analyzed by *matrix-assisted laser desorption ionization* (MALDI) as illustrated in Figure 25.10.

MALDI offers two main advantages over traditional mass spectrometric methods.

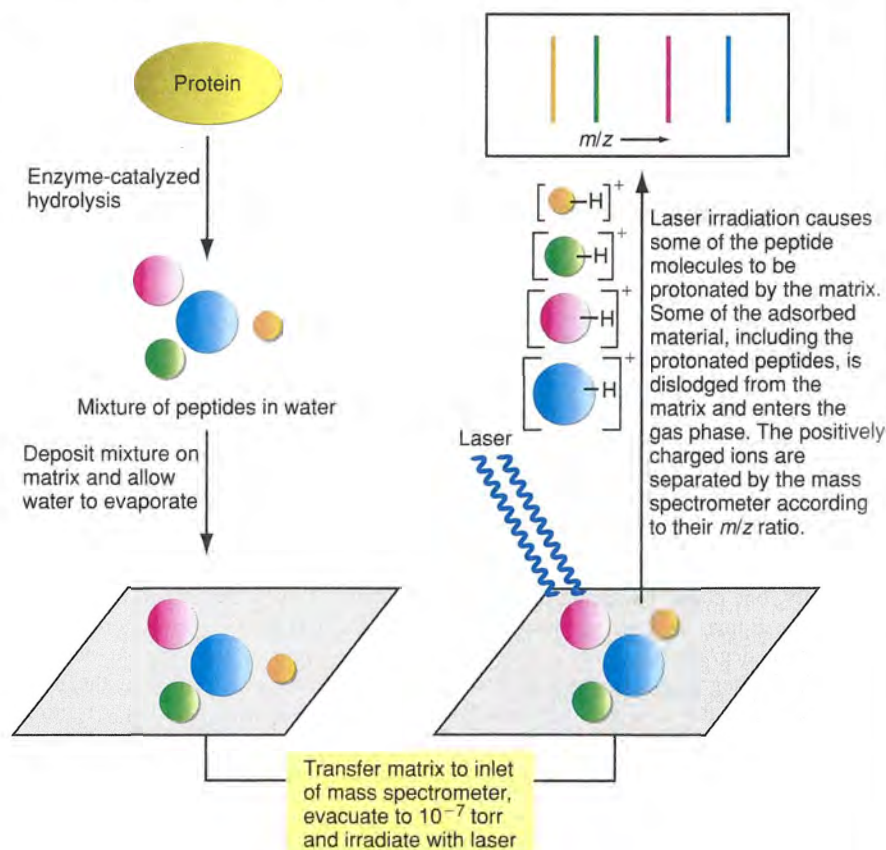
1. Substances, such as peptides, that lack sufficient volatility to be vaporized for analysis by conventional mass spectrometry can be vaporized by MALDI.
2. The species analyzed by the mass spectrometer is the conjugate acid of a peptide (peptide +  $H^+$ ). Unlike the highly energetic ions generated by electron impact (see Section 13.24), the cations produced by MALDI have little tendency to fragment. Consequently, a mixture of peptide fragments gives a mass

spectrum dominated by peaks with  $m/z$  values corresponding to those of the individual protonated peptides.

With the aid of freely available Internet tools and databases, the MALDI data set is compared with known proteins to generate a list of potential matches. The analyst inputs the peptidase used to digest the original protein and the  $m/z$  values of the peptides displayed in the MALDI spectrum. As specified by the search criteria, the search delivers (in a matter seconds!) a list of peptide sequences and the proteins these sequences contain. Next, a different peptidase is used to hydrolyze the protein, to provide a second set of peptides that are also analyzed by MALDI and matched against the database. MALDI mass spectrometry compares the amino acid *composition* of unsequenced peptides with the amino acid *sequence* of known proteins in order to identify an unknown protein. The procedure is repeated until the list of potential matches is narrowed to a single known protein, additional data are needed, or it becomes likely that the protein is new.

**Figure 25.10**

The molecular weights of all of the peptides in a mixture obtained by the enzyme-catalyzed hydrolysis of a protein are determined simultaneously by mass spectrometry using matrix-assisted laser desorption ionization (MALDI).

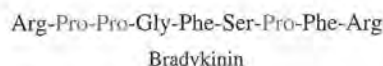


### Problem 25.20

Which two of the standard amino acids cannot be differentiated on the basis of their  $m/z$  ratio?

## 25.14 The Strategy of Peptide Synthesis

One way to confirm the structure proposed for a peptide is to synthesize a peptide having a specific sequence of amino acids and compare the two. This was done, for example, in the case of *bradykinin*, a peptide present in blood that acts to lower blood pressure. Excess bradykinin, formed as a response to the sting of wasps and other insects containing substances in their venom that stimulate bradykinin release, causes severe local pain. Bradykinin was originally believed to be an octapeptide containing two proline residues; however, a nonapeptide containing three prolines in the following sequence was synthesized and determined to be identical with natural bradykinin in every respect, including biological activity:



A reevaluation of the original sequence data established that natural bradykinin was indeed the nonapeptide shown. Here the synthesis of a peptide did more than confirm structure; synthesis was instrumental in determining structure.

The synthesis of peptides is an important area of drug development. For many years, the high cost of peptide synthesis and the rapid degradation of peptides when administered in the body hampered the progress of peptide drug discovery. Recent advances in peptide manufacturing and drug delivery have increased the use of peptide-based pharmaceuticals. Two of the most prominent peptide drugs are insulin and calcitonin.

A promising method of drug delivery uses liposomes (see Section 24.4).

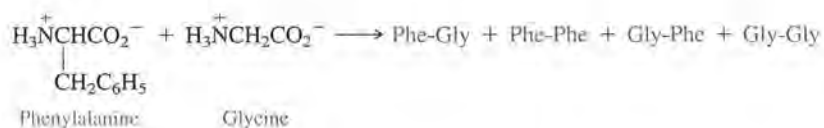
The insulin required for the treatment of diabetes used to be obtained by extraction from the pancreas glands of cows and pigs. Since the early 1980s, this “natural” insulin has been replaced by “synthetic” human insulin prepared by recombinant DNA technology (Section 26.17). Synthetic insulin is not only identical to human insulin, it is both safer and less expensive than insulin obtained from animals. A somewhat smaller polypeptide, *calcitonin* with 32 amino acids, is prepared by more traditional methods of synthetic organic chemistry. Synthetic calcitonin is identical to that obtained from salmon and is widely used for the treatment of osteoporosis. How calcitonin acts remains uncertain, but one possibility is that it maintains bone mass, not by increasing the rate of bone growth, but by decreasing the rate of bone loss. The biological receptor for calcitonin is located on a type of bone cell known as an osteoclast.

Other than the biochemical methods typified by the synthesis of insulin, there are two major approaches to peptide synthesis:

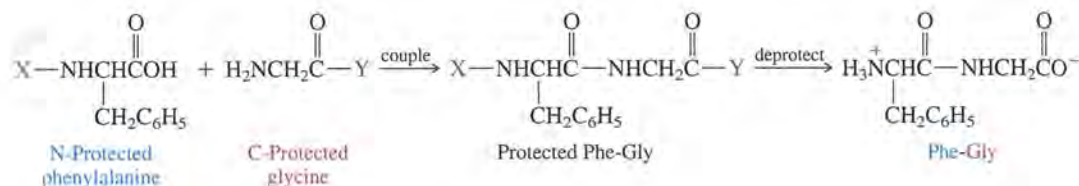
1. Solution phase
2. Solid phase

Although the two approaches differ in respect to the phase in which the synthesis is carried out, the overall strategy is the same in both.

The objective in peptide synthesis may be simply stated: to connect amino acids in a prescribed sequence by amide bond formation between them. A number of very effective methods and reagents have been designed for peptide bond formation, so that the joining together of amino acids by amide linkages is not difficult. The real difficulty lies in ensuring that the correct sequence is obtained. This can be illustrated by considering the synthesis of a representative dipeptide, Phe-Gly. Random peptide bond formation in a mixture containing phenylalanine and glycine would be expected to lead to four dipeptides:



To direct the synthesis so that only Phe-Gly is formed, the amino group of phenylalanine and the carboxyl group of glycine must be protected so that they cannot react under the conditions of peptide bond formation. We can represent the peptide bond formation step by the following equation, where X and Y are amine- and carboxyl-protecting groups, respectively:



Thus, the synthesis of a dipeptide of prescribed sequence requires at least three operations:

1. *Protect* the amino group of the N-terminal amino acid and the carboxyl group of the C-terminal amino acid.
2. *Couple* the two protected amino acids by amide bond formation between them.
3. *Deprotect* the amino group at the N terminus and the carboxyl group at the C terminus.

Higher peptides are prepared in an analogous way by a direct extension of the logic just outlined for the synthesis of dipeptides.

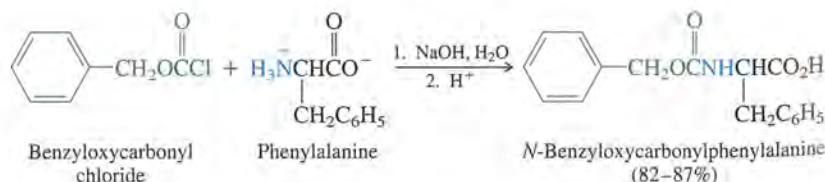
Sections 25.15 through 25.17 describe the chemistry associated with the protection and deprotection of amino and carboxyl functions, along with methods for peptide bond formation. The focus in those sections is on solution-phase peptide synthesis. Section 25.18 shows how these methods are adapted to solid-phase synthesis.

## 25.15 Amino Group Protection

The reactivity of an amino group is suppressed by converting it to an amide, and amino groups are most often protected by acylation. The benzyloxycarbonyl group

( $\text{C}_6\text{H}_5\text{CH}_2\text{OC}-$ ) is one of the most often used amino-protecting groups. It is attached by acylation of an amino acid with benzyloxycarbonyl chloride.

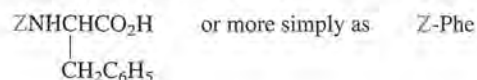
Another name for the benzyloxycarbonyl group is *carbobenzyloxy*. This name, and its abbreviation Cbz, are often found in the older literature, but are no longer a part of IUPAC nomenclature.



### Problem 25.21

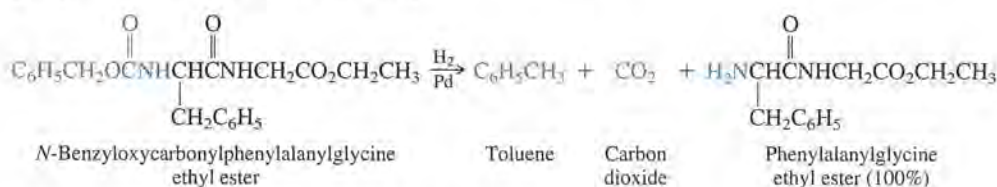
Lysine reacts with two equivalents of benzyloxycarbonyl chloride to give a derivative containing two benzyloxycarbonyl groups. What is the structure of this compound?

Just as it is customary to identify individual amino acids by abbreviations, so too with protected amino acids. The approved abbreviation for a benzyloxycarbonyl group is the letter Z. Thus, *N*-benzyloxycarbonylphenylalanine is represented as

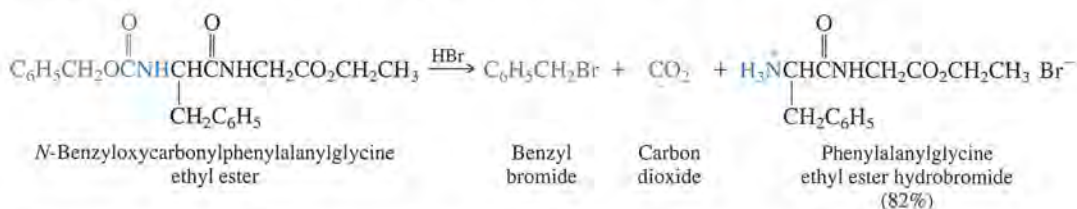


The value of the benzyloxycarbonyl protecting group is that it is easily removed by reactions other than hydrolysis. In peptide synthesis, amide bonds are formed. We protect the N terminus as an amide but need to remove the protecting group without cleaving the very amide bonds we labored so hard to construct. Removing the protecting group by hydrolysis would surely bring about cleavage of peptide bonds as well. One advantage that the benzyloxycarbonyl protecting group enjoys over more familiar acyl groups such as acetyl is that it can be removed by *hydrogenolysis* in the presence of palladium. The following equation illustrates this for the removal of the benzyloxycarbonyl protecting group from the ethyl ester of Z-Phe-Gly:

*Hydrogenolysis* refers to the cleavage of a molecule under conditions of catalytic hydrogenation.

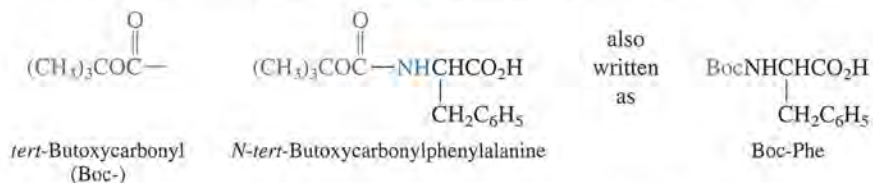


Alternatively, the benzyloxycarbonyl protecting group may be removed by treatment with hydrogen bromide in acetic acid:

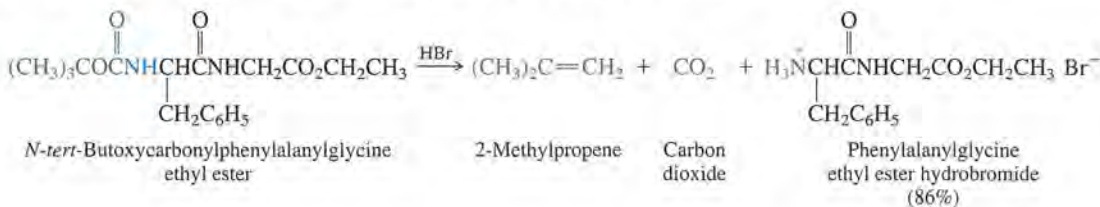


Deprotection by this method rests on the ease with which benzyl esters are cleaved by nucleophilic attack at the benzylic carbon in the presence of strong acids. Bromide ion is the nucleophile.

A related N-terminal-protecting group is *tert*-butoxycarbonyl, abbreviated *Boc*:

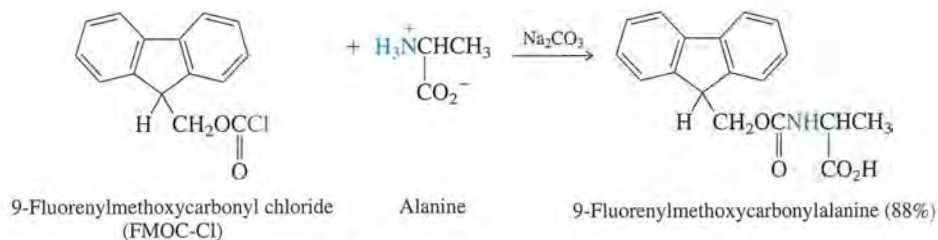


Like the benzyloxycarbonyl protecting group, the Boc group may be removed by treatment with hydrogen bromide (it is stable to hydrogenolysis, however):

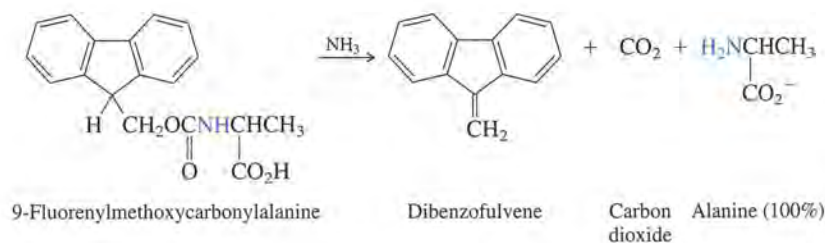


The *tert*-butyl group is cleaved as the corresponding carbocation. Loss of a proton from *tert*-butyl cation converts it to 2-methylpropene. Because of the ease with which a *tert*-butyl group is cleaved as a carbocation, other acidic reagents, such as trifluoroacetic acid, may also be used.

A third N-protecting group option is 9-fluorenylmethoxycarbonyl (Fmoc).

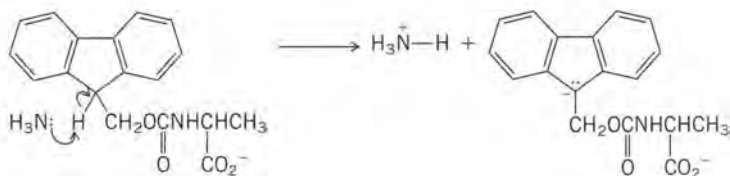


Fmoc differs from Z and Boc in that it is removed under basic conditions.

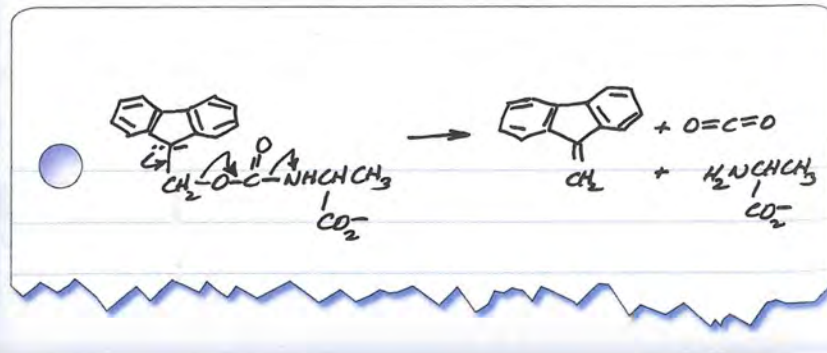


### Problem 25.22

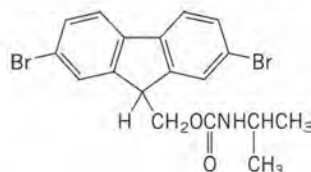
The mechanism of removing the Fmoc protecting group rests on the fact its five-membered ring is structurally analogous to 1,3-cyclopentadiene, which makes its hydrogen relatively acidic. Ammonia removes this hydrogen to give an anion:



(a) Use curved arrows to show how the anion produced in this step dissociates.

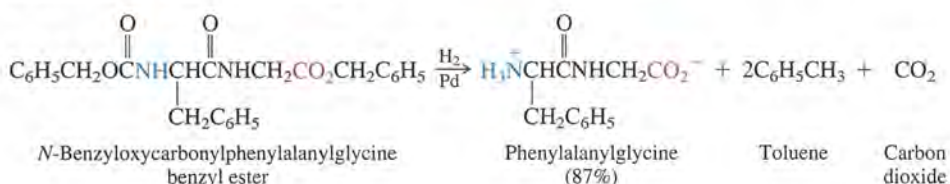


- (b) The dibromo analog of the Fmoc protecting group can be removed with pyridine, which is a weaker base than is used to remove the Fmoc group. Explain why the dibromo derivative is more easily removed by base than Fmoc.



## 25.16 Carboxyl Group Protection

Carboxyl groups of amino acids and peptides are normally protected as esters. Methyl and ethyl esters are prepared by Fischer esterification. Deprotection of methyl and ethyl esters is accomplished by hydrolysis in base. Benzyl esters are a popular choice because they can also be removed by hydrogenolysis. Thus a synthetic peptide, protected at its N terminus with a Z group and at its C terminus as a benzyl ester, can be completely deprotected in a single operation.

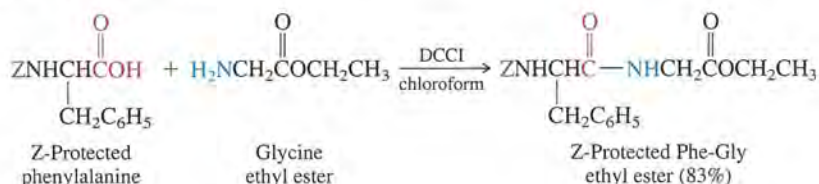


Several of the amino acids listed in Table 25.1 bear side-chain functional groups, which must also be protected during peptide synthesis. In most cases, protecting groups are available that can be removed by hydrogenolysis.

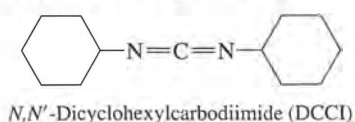
## 25.17 Peptide Bond Formation

To form a peptide bond between two suitably protected amino acids, the free carboxyl group of one of them must be *activated* so that it is a reactive acylating agent. The most familiar acylating agents are acyl chlorides, and they were once extensively used to couple amino acids. Certain drawbacks to this approach, however, led chemists to seek alternative methods.

In one method, treatment of a solution containing the N-protected and the C-protected amino acids with *N,N'*-dicyclohexylcarbodiimide (DCCI) leads directly to peptide bond formation:



*N,N'*-Dicyclohexylcarbodiimide has the structure:



Mechanism 25.4 shows how DCCl promotes the formation of the peptide bond shown in the preceding equation.

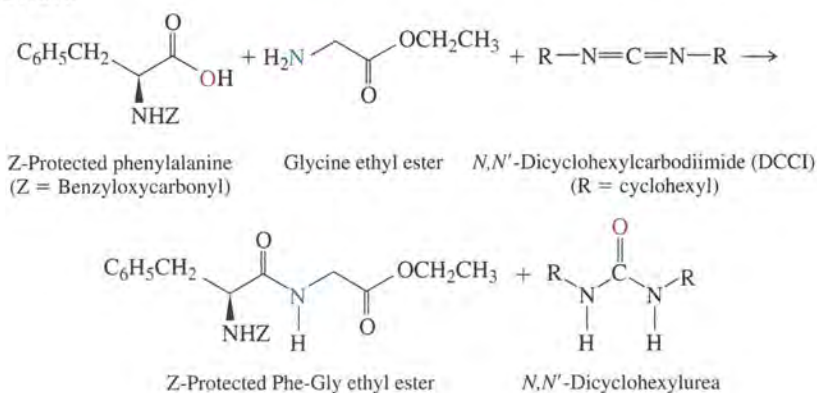
### Problem 25.23

Show the steps involved in the synthesis of Ala-Leu from alanine and leucine using benzyloxycarbonyl and benzyl ester protecting groups and DCCl-promoted peptide bond formation.

### Mechanism 25.4

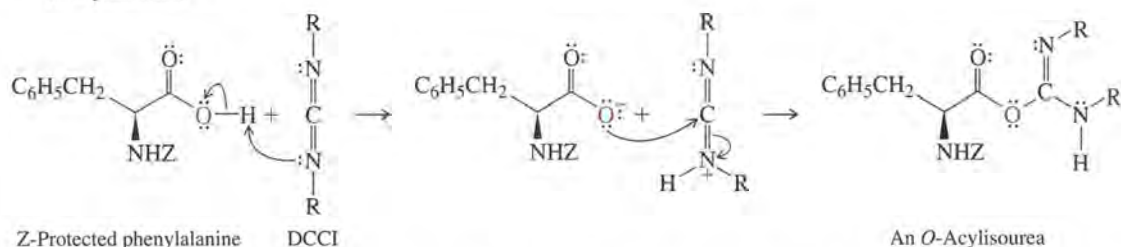
#### Amide Bond Formation Between a Carboxylic Acid and an Amine Using *N,N'*-Dicyclohexylcarbodiimide

##### THE OVERALL REACTION:

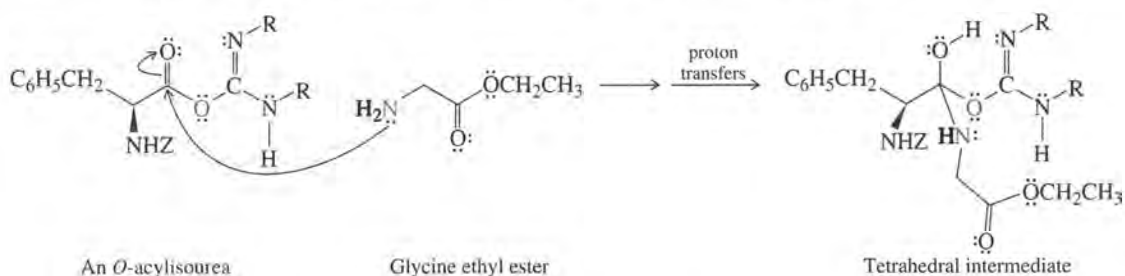


##### THE MECHANISM:

**Step 1:** In the first stage of the reaction, the carboxylic acid adds to one of the double bonds of DCCI to give an *O*-acylisourea.

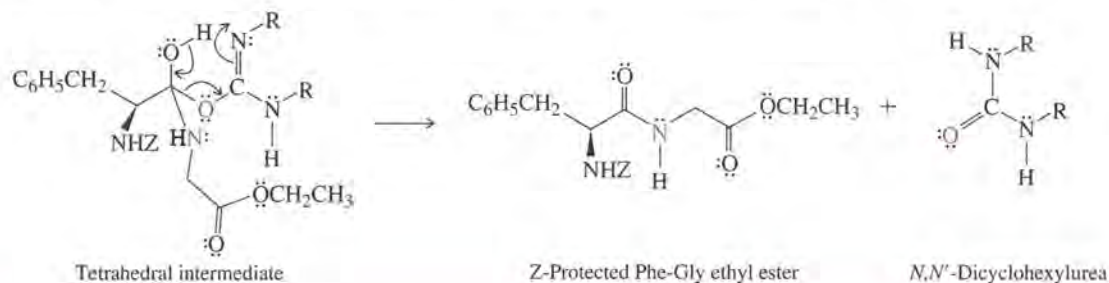


**Step 2:** Structurally, *O*-acylisoureas resemble acid anhydrides and are powerful acylating agents. In the reaction's second stage the amine adds to the carbonyl group of the *O*-acylisourea to give a tetrahedral intermediate.

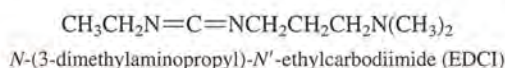


*Continued*

**Step 3:** The tetrahedral intermediate dissociates to an amide and *N,N'*-dicyclohexylurea.



The *N,N'*-dicyclohexylurea that is produced in peptide couplings with DCCI is soluble in the same solvents as the product, making it more difficult to separate the by-products during purification. An alternative carbodiimide that is used in peptide synthesis is *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide, known as EDCI. Both EDCI and the urea that is produced from it are water-soluble and can be more easily separated from the peptide product. EDCI can also be used in organic solvents.



### Problem 25.24

What is the structure of the urea that is produced in the coupling reaction of a protected amino acid and a carboxylic acid with EDCI?

Higher peptides are prepared either by stepwise extension of peptide chains, one amino acid at a time, or by coupling of fragments containing several residues (the *fragment condensation* approach). Human pituitary adrenocorticotrophic hormone (ACTH), for example, has 39 amino acids and was synthesized by coupling of smaller peptides containing residues 1–10, 11–16, 17–24, and 25–39. An attractive feature of this approach is that the various protected peptide fragments may be individually purified, which simplifies the purification of the final product. Among the substances that have been synthesized by fragment condensation are insulin (51 amino acids) and the protein ribonuclease A (124 amino acids). In the stepwise extension approach, the starting peptide in a particular step differs from the coupling product by only one amino acid residue and the properties of the two peptides may be so similar as to make purification by conventional techniques all but impossible. The solid-phase method described in the following section overcomes many of the difficulties involved in the purification of intermediates.

## 25.18 Solid-Phase Peptide Synthesis: The Merrifield Method

In 1962, R. Bruce Merrifield of Rockefeller University reported the synthesis of the nonapeptide bradykinin by a novel method. In Merrifield's method, peptide coupling and deprotection are carried out not in homogeneous solution but at the surface of an insoluble polymer, or *solid support*. Beads of a copolymer prepared from styrene containing about 2% divinylbenzene are treated with chloromethyl methyl ether and tin(IV) chloride to give a resin in which about 10% of the aromatic rings bear —CH<sub>2</sub>Cl groups (Figure 25.11). The growing peptide is anchored to this polymer, and excess reagents, impurities, and by-products are removed by thorough washing after each operation. This greatly simplifies the purification of intermediates.

Merrifield was awarded the 1984 Nobel Prize in Chemistry for developing the solid-phase method of peptide synthesis.

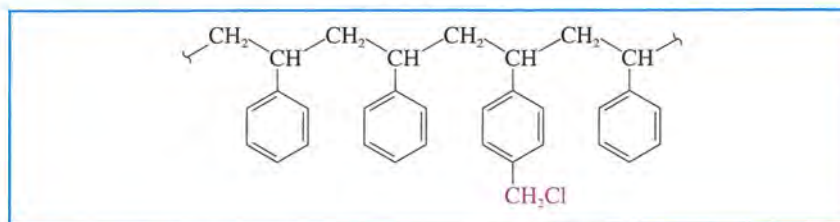


Figure 25.11

A section of polystyrene showing one of the benzene rings modified by chloromethylation. Individual polystyrene chains in the resin used in solid-phase peptide synthesis are connected to one another at various points (cross-linked) by adding a small amount of *p*-divinylbenzene to the styrene monomer. The chloromethylation step is carried out under conditions such that only about 10% of the benzene rings bear  $\text{—CH}_2\text{Cl}$  groups.

The Merrifield procedure has been adapted to accommodate Fmoc as well as Boc protecting groups.

The actual process of **solid-phase peptide synthesis**, outlined in Figure 25.12, begins with the attachment of the C-terminal amino acid to the chloromethylated polymer in step 1. Nucleophilic substitution by the carboxylate anion of an *N*-Boc-protected C-terminal amino acid displaces chloride from the chloromethyl group of the polymer to form an ester, protecting the C terminus while anchoring it to a solid support. Next, the Boc group is removed by treatment with acid (step 2), and the polymer containing the unmasked N terminus is washed with a series of organic solvents. By-products are removed, and only the polymer and its attached C-terminal amino acid residue remain. Next (step 3), a peptide bond to an *N*-Boc-protected amino acid is formed by condensation in the presence of *N,N'*-dicyclohexylcarbodiimide. Again, the polymer is washed thoroughly. The Boc protecting group is then removed by acid treatment (step 4), and after washing, the polymer is now ready for the addition of another amino acid residue by a repetition of the cycle. When all the amino acids have been added, the synthetic peptide is removed from the polymeric support by treatment with hydrogen bromide in trifluoroacetic acid.

By successively adding amino acid residues to the C-terminal amino acid, it took Merrifield only eight days to synthesize the nonapeptide bradykinin in 68% yield. The biological activity of synthetic bradykinin was identical with that of natural material.

#### Problem 25.25

Starting with phenylalanine and glycine, outline the steps in the preparation of Phe-Gly by the Merrifield method.

Merrifield successfully automated all the steps in solid-phase peptide synthesis, and computer-controlled equipment is commercially available to perform this synthesis. Using an early version of his “peptide synthesizer,” in collaboration with coworker Bernd Gutte, Merrifield reported the synthesis of the enzyme ribonuclease in 1969. It took them only six weeks to perform the 369 reactions and 11,391 steps necessary to assemble the sequence of 124 amino acids of ribonuclease.

Solid-phase peptide synthesis does not solve all purification problems, however. Even if every coupling step in the ribonuclease synthesis proceeded in 99% yield, the product would be contaminated with many different peptides containing 123 amino acids, 122 amino acids, and so on. Thus, Merrifield and Gutte’s six weeks of synthesis was followed by four months spent in purifying the final product. The technique has since been refined to the point that yields at the 99% level and greater are achieved with current instrumentation, and thousands of peptides and peptide analogs have been prepared by the solid-phase method.

Merrifield’s concept of a solid-phase method for peptide synthesis and his development of methods for carrying it out set the stage for an entirely new way to do chemical reactions. Solid-phase synthesis has been extended to include numerous other classes of

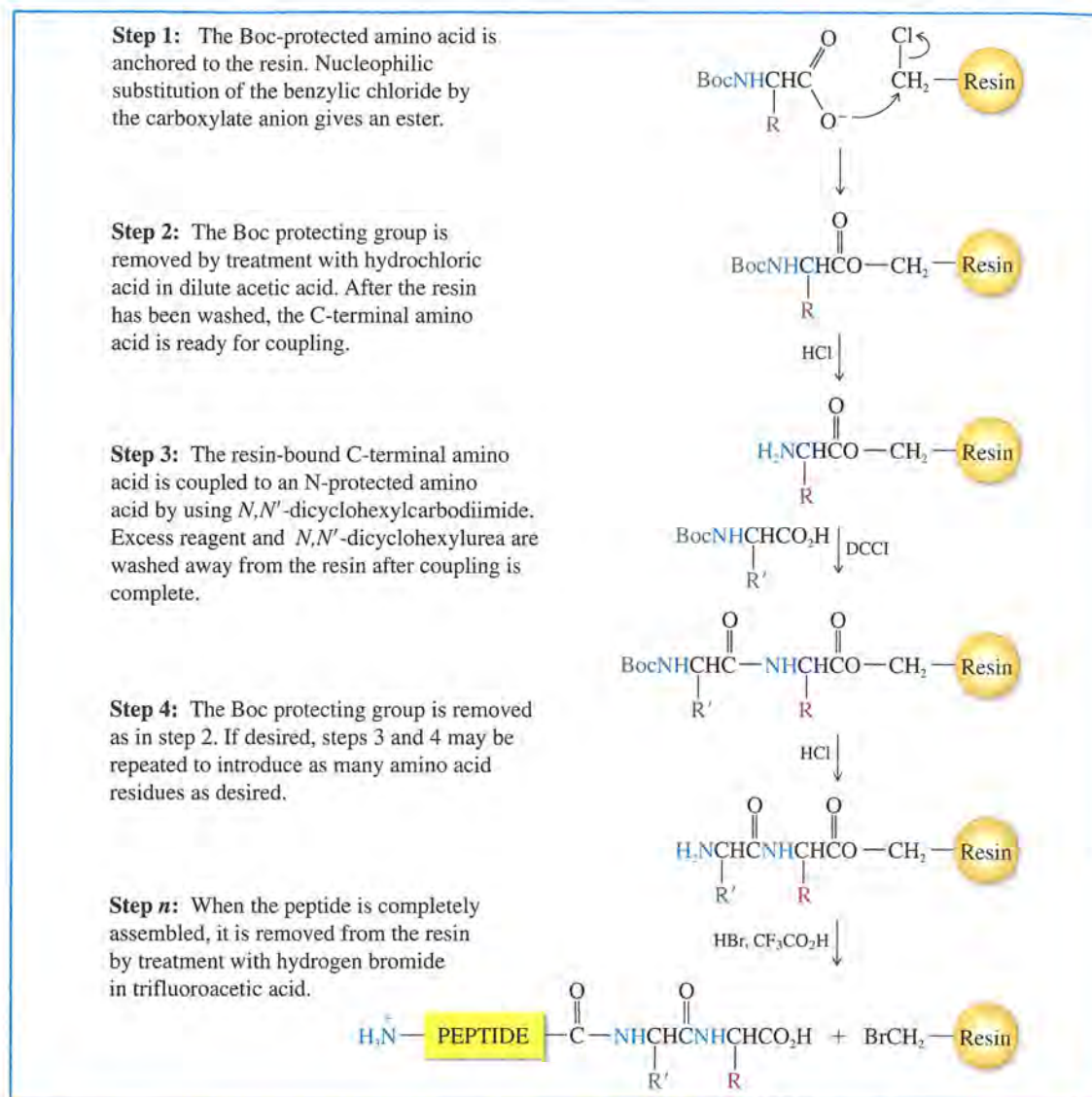


Figure 25.12

Peptide synthesis by the solid-phase method. Amino acid residues are attached sequentially beginning at the C terminus.

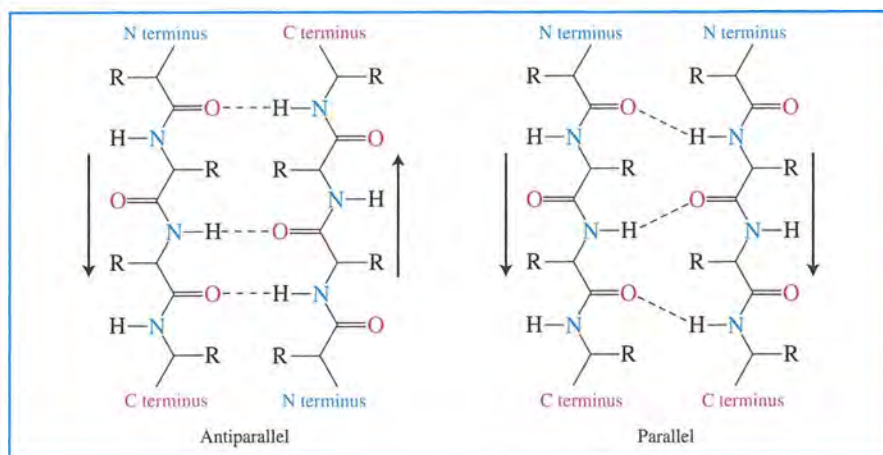
compounds and has helped spawn a whole new field called **combinatorial chemistry**. Combinatorial synthesis allows a chemist, using solid-phase techniques, to prepare hundreds of related compounds (called *libraries*) at a time.

## 25.19 Secondary Structures of Peptides and Proteins

The primary structure of a peptide is its amino acid sequence. The **secondary structure** is the conformational relationship of nearest neighbor amino acids with respect to each other. On the basis of X-ray crystallographic studies and careful examination of molecular models, Linus Pauling and Robert B. Corey of the California Institute of Technology showed that certain peptide conformations were more stable than others. Two arrangements, the

**Figure 25.13**

Hydrogen bonding between the carbonyl oxygen of one peptide chain and the amide N—H of another in a  $\beta$ -pleated sheet. In the antiparallel arrangement, the N-terminus  $\rightarrow$  C-terminus direction of one chain is opposite to that of the other. In the parallel arrangement, the N-terminus  $\rightarrow$  C-terminus direction is the same for both chains.



$\alpha$  helix and the  $\beta$  sheet, stand out as secondary structural units that are both particularly stable and commonly encountered. Both of these incorporate two important features:

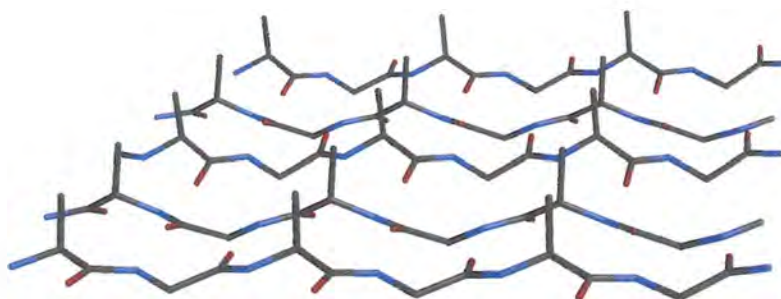
1. The geometry of the peptide bond is planar and the main chain is arranged in an anti conformation (see Section 25.7).
2. Hydrogen bonding can occur when the N—H group of one amino acid unit and the C=O group of another are close in space; conformations that maximize the number of these hydrogen bonds are stabilized by them.

Chains in a  $\beta$  sheet exist in an extended conformation with hydrogen bonds between a carbonyl oxygen of one chain and an amide N—H of another (Figure 25.13). Both the parallel and antiparallel arrangements of chains occur in proteins. Some of the space between peptide chains is occupied by the amino acid side chains, represented by R in Figure 25.13. Van der Waals repulsive forces involving these substituents cause the chains to rotate with respect to one another, giving a rippled effect known as a  *$\beta$ -pleated sheet* (Figure 25.14).

The  $\beta$ -pleated sheet is an important secondary structure in proteins that are rich in amino acids with small side chains such as H (glycine),  $\text{CH}_3$  (alanine), and  $\text{CH}_2\text{OH}$  (serine). The model in Figure 25.14 is a portion of the calculated structure for a sheet composed of antiparallel strands containing only glycine and alanine in alternating order (Gly-Ala-Gly-Ala-, etc.). It was designed to resemble *fibroin*, the major protein of silk. Fibroin is almost entirely pleated sheet, and over 80% of it is a repeating sequence of the six-residue unit -Gly-Ser-Gly-Ala-Gly-Ala-. Because the polypeptide backbone adopts an extended zigzag conformation, silk, unlike wool for example, resists stretching.

**Figure 25.14**

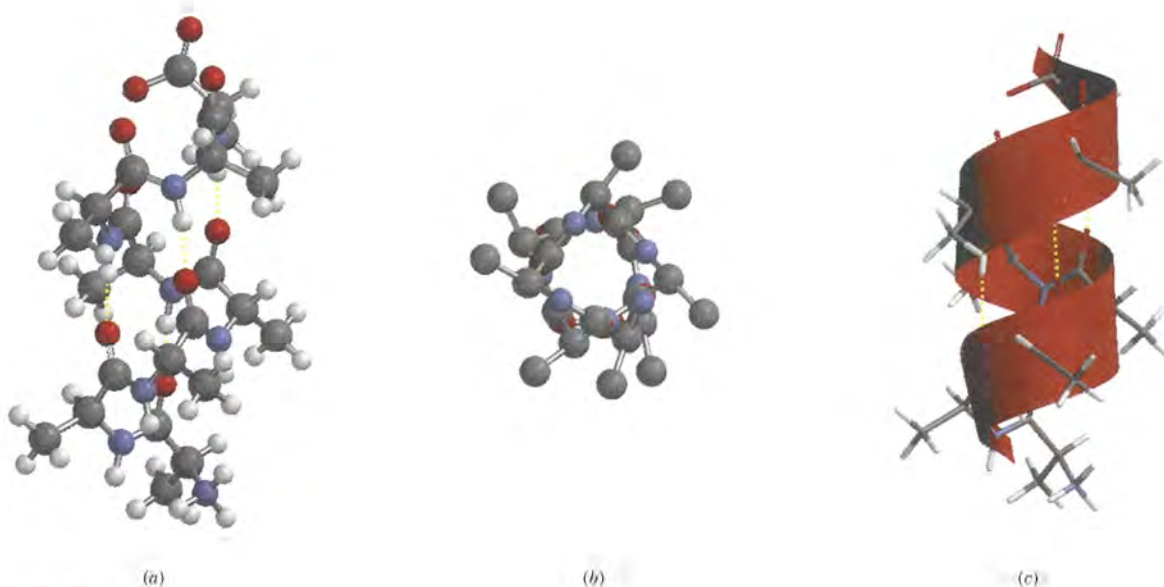
The  $\beta$ -pleated sheet secondary structure of a protein, composed of alternating glycine and alanine residues.



**Problem 25.26**

The methyl groups of the alanine residues of the  $\beta$  sheet in Figure 25.14 all point upward. If this pleated sheet were composed of only alanine residues instead of being Gly-Ala-Gly-Ala, etc. what would be the pattern of methyl groups? Would they all point up, alternate up and down, or be random?

The  $\alpha$  helix is another commonly encountered secondary structure. Figure 25.15 gives three views of an  $\alpha$ -helix model constructed from eight L-alanine residues. Part *a* of the figure is a ball-and-spoke model; part *b* is a view through the center of the helix along its axis. The helix is right-handed with about 3.6 amino acids per turn and is stabilized by hydrogen bonds between the carbonyl oxygens and N—H protons. View *b* shows how the methyl groups of L-alanine project outward from the main chain. This outward orientation of amino acid side chains makes them the points of contact with other amino acids of the same chain, with different protein chains, and with other biomolecules. Part *c* of the figure uses a ribbon to trace the peptide backbone. The ribbon helps distinguish front from back, makes the right-handedness of the helix more apparent, and is especially useful when looking at how proteins are folded.

**Figure 25.15**

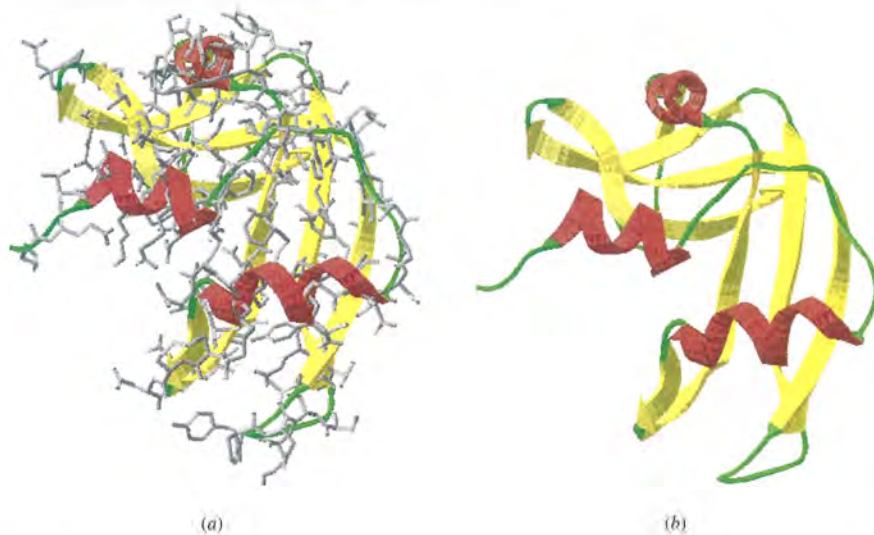
Molecular models of an  $\alpha$  helix composed of eight alanine residues. The N terminus is at the bottom. (a) A ball-and-spoke model. Hydrogen bonds are shown as dashed lines. (b) The same model looking up the helical axis from the bottom. Hydrogens have been omitted for clarity. The helix is right-handed, and all of the methyl groups point outward. (c) A tube model framed in a ribbon that traces the path of the helix.

The protein components of muscle (*myosin*) and wool ( $\alpha$ -*keratin*) contain high percentages of  $\alpha$  helix. When wool is stretched, hydrogen bonds break and the peptide chain is elongated. Covalent S—S bonds between L-cysteine residues limit the extent to which the chain can be stretched, however, and once the stretching force is removed the hydrogen bonds re-form spontaneously.

Most proteins cannot be described in terms of a single secondary structure. Rather, most are mixtures of  $\alpha$  helix and  $\beta$  sheet, interspersed with regions of **random coils** that have no regular pattern. Figure 25.16 shows a model of ribonuclease, an enzyme that catalyzes the hydrolysis of RNA. The helical regions are shown in red, the  $\beta$  sheets in yellow. Of the 124 amino acids in this protein, 24 are represented in three sections of

Figure 25.16

Molecular models of ribonuclease. Red ribbons identify sequences where the secondary structure is a helix; yellow ribbons identify strands of  $\beta$  sheet. Arrowheads point in the direction from the N terminus to the C terminus. (a) Shows both a molecular model and ribbons. (b) Shows only the ribbons.



$\alpha$  helix. There are two  $\beta$  sheets, one with three strands accounting for 21 amino acids, the other with four strands and 20 amino acids. The strands of each  $\beta$  sheet belong to the same chain and are brought within hydrogen bonding distance because of how the chain is folded. Indeed, the formation of hydrogen bonds such as these is one of the factors that contributes to chain folding.

Another protein which has regions of  $\alpha$  helix and  $\beta$  sheet is green fluorescent protein, or GFP. In GFP, the  $\beta$  sheet structure is barrel-shaped, and the  $\alpha$  helix runs along the center of the barrel, where it is shielded from the exterior (Figure 25.17). The  $\alpha$  helix in GFP contains an arrangement of amino acids that undergoes fluorescence. Through the use of recombinant DNA technology (Section 26.17), GFP can be attached to other proteins that are normally invisible, but now can be imaged using fluorescence microscopy. Using GFP as a biomarker, it is possible to monitor the roles of different proteins in the body, in part because GFP has very low toxicity toward living cells. The development of nerve cells in the brain, the spread of

Figure 25.17

Barrel-shaped green fluorescent protein (GFP) has an outer  $\beta$ -sheet structure and an  $\alpha$  helix in the inner region.



cancer cells, and the damage to neurons that occurs during Alzheimer's disease can be followed by techniques that use GFP. The 2008 Nobel Prize in Chemistry was awarded to Osamu Shimomura, Martin Chalfie, and Roger Tsien for their work in the development of GFP.

## 25.20 Tertiary Structure of Polypeptides and Proteins

The way a protein chain is folded, its **tertiary structure**, affects both its physical properties and its biological function. The two main categories of protein tertiary structure are **fibrous** and **globular**.

1. Fibrous proteins are bundles of elongated filaments of protein chains and are insoluble in water.
2. Globular proteins are approximately spherical and are either soluble or form colloidal dispersions in water.

The primary structure of a protein, its amino acid sequence, is the main determinant of tertiary structure. Secondary structure also contributes by limiting the number of conformations available to a polypeptide chain.

Fibrous proteins, being insoluble in water often have a structural or protective function. The most familiar fibrous proteins are the keratins and collagen.  $\alpha$ -Keratin (Figure 25.18) is based on the  $\alpha$ -helix secondary structure and is the protein structural component of hair, wool, nails, claws, quills, horns, and the outer layer of skin.  $\beta$ -Keratin is based on the  $\beta$ -sheet secondary structure and occurs in silk as fibroin. L-Cysteine is especially abundant in keratins, where it can account for more than 20% of the amino acids present. Collagen occurs mainly in connective tissue (cartilage and tendons) and has a triple helix structure.

Globular proteins include most enzymes and function in aqueous environments. About 65% of the mass of most cells, for example, is water. When placed in water, nonpolar materials, including nonpolar amino acid side chains, cause nearby water molecules to adopt a more ordered arrangement, reducing the entropy of water. This is called the **hydrophobic effect**. The unfavorable negative  $\Delta S$  is moderated if the protein adopts a spherical shape which places nonpolar side chains inside and polar ones on the surface. Of the various globular arrangements, the one that best offsets the entropy loss with attractive forces among the side chains is the tertiary structure adopted by the protein in its normal, or *native state*.

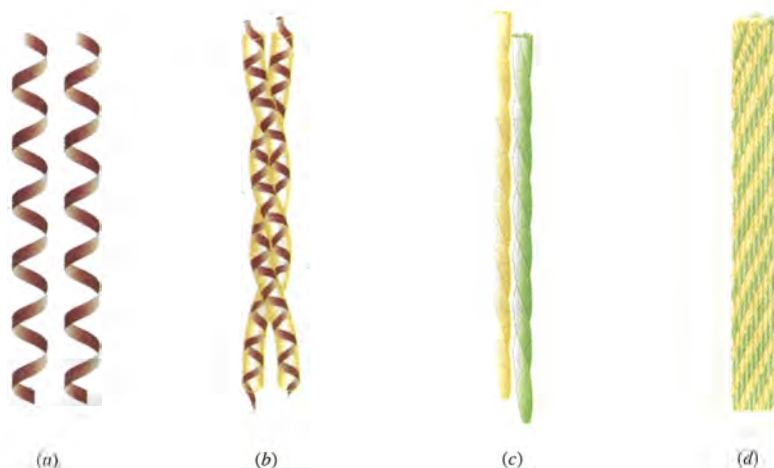


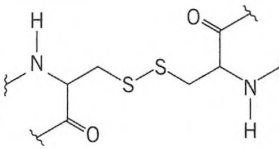
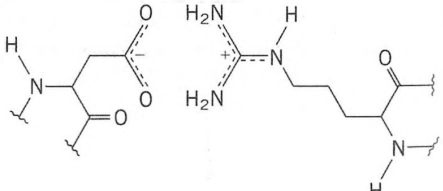
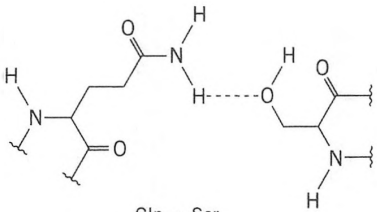
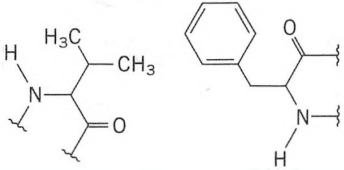
**Figure 25.18**

$\alpha$ -Keratin. Two  $\alpha$  helices (a) combine to give a coiled coil (b). A pair of coiled coils is a protofilament (c). Four protofilaments give a filament (d), which is the structural material from which the fibrous protein is assembled.

Table 25.4 lists the attractive forces that most influence protein tertiary structure. The strongest of these is the covalent S—S bond that unites two cysteine residues. This *disulfide bridge* can form between the —CH<sub>2</sub>SH groups of two cysteines, which, although they may be remote from each other in respect to the amino acid sequence, become neighbors when the chain is folded. Formation of the disulfide bond connecting the two stabilizes the local folded arrangement. A typical globular protein normally has only a small number of disulfide bridges. Of the 124 amino acids in ribonuclease (see Figure 25.16 in Section 25.19), 6 are cysteines and each participates in a disulfide bridge; one bridge unites Cys-26 and Cys-84, another Cys-58 and Cys-110, and a third Cys-65 and Cys-72.

The noncovalent interactions are all much weaker than the S—S covalent bond. Among them, the electrostatic attraction between positively and negatively charged side chains, called a salt bridge, is the strongest, followed by hydrogen bonding, then van der Waals forces. Keep in mind though, that the total contribution of the various forces depends not only on the magnitude of an interaction but also on their number.

**TABLE 25.4 Covalent and Noncovalent Interactions Between Amino Acid Side Chains in Proteins**

| Description        | Type of interaction                                                                        | Example                                                                                               |
|--------------------|--------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| <b>Covalent</b>    |                                                                                            |                                                                                                       |
| Disulfide bridge   | S—S bond between two cysteines                                                             |  <p>CysS—SCys</p>  |
| <b>Noncovalent</b> |                                                                                            |                                                                                                       |
| Salt bridge        | Electrostatic attraction between oppositely charged ions                                   |  <p>Asp---Arg</p> |
| Hydrogen bond      | Positively polarized H of O—H or N—H group interacts with an electronegative atom (O or N) |  <p>Gln---Ser</p> |
| Van der Waals      | Induced-dipole/induced-dipole attraction (dispersion force) between nonpolar side chains   |  <p>Val---Phe</p> |

Disulfide bridges may be strong, but there are usually only a few of them. Van der Waals forces are weak, but they outnumber all the other intermolecular attractive forces.

### Problem 25.27

Table 25.4 shows a salt bridge between aspartic acid and arginine. Sketch the analogous electrostatic attraction between lysine and an amino acid other than aspartic acid from Table 25.1.

Knowing how the protein chain is folded is a key element in understanding how an **enzyme** catalyzes a reaction. Biochemical processes are usually related to the core reaction types of organic chemistry and involve similar key intermediates. The reactions, however, are much faster and more selective. In proposing an enzyme-catalyzed mechanism for a reaction such as amide or ester hydrolysis, it is customary to assume it proceeds by way of a tetrahedral intermediate, then modify the usual nucleophilic acyl substitution mechanism by assigning various catalytic functions to selected amino acid side chains of the enzyme.

We saw in Section 25.10 that carboxypeptidase A catalyzes the hydrolysis of the peptide bond to the C-terminal amino acid of polypeptides. Carboxypeptidase A is a *metalloprotein*; it contains a  $\text{Zn}^{2+}$  ion, which is essential for catalytic activity. The X-ray crystal structure of carboxypeptidase A (Figure 25.19) locates this  $\text{Zn}^{2+}$  ion in a hydrophobic cavity near the center of the enzyme, where it is held by coordination to a glutamic acid residue (Glu-72) and two histidines (His-69 and His-196). This is the same region, called the **active site**, where the substrate binds. The substrate in the case of carboxypeptidase is a peptide, especially a peptide with a hydrophobic C-terminal amino acid such as phenylalanine or tyrosine. In addition to being hydrophobic, as is the active site, the substrate is bound by an electrostatic attraction between its negatively charged carboxylate and the positively charged side chain of Arg-145. Mechanism 25.5 shows the interactions of the side chains of carboxypeptidase A with  $\text{Zn}^{2+}$  and a peptide, then describes the mechanism for cleaving the peptide bond to the terminal amino acid. Side chains other than those shown in Mechanism 25.5 have been implicated but have been omitted. The main feature of the mechanism is its relationship to the mechanism of nucleophilic acyl substitution (on which it was patterned). Not only does the enzyme bring the substrate and catalytically active functions together at the active site, but by stabilizing the tetrahedral intermediate it lowers the activation energy for its formation and increases the reaction rate.

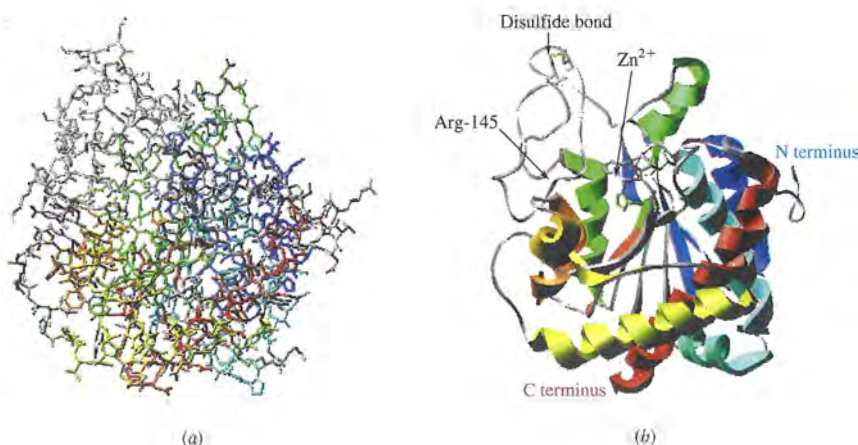


**Figure 25.19**

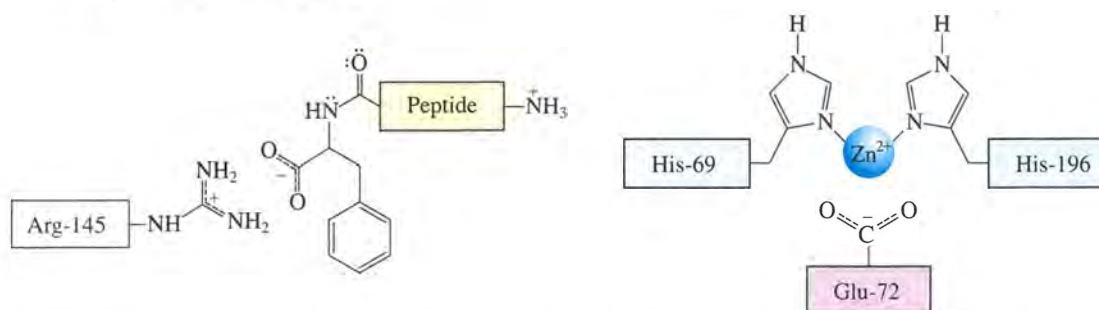
The structure of carboxypeptidase A displayed as (a) a tube model and (b) a ribbon diagram. The most evident feature illustrated by (a) is the globular shape of the enzyme. The ribbon diagram emphasizes the folding of the chain.

## Mechanism 25.5

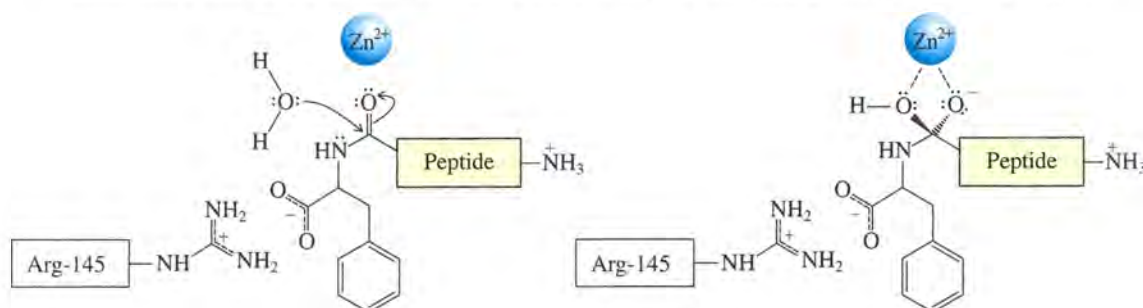
## Carboxypeptidase-Catalyzed Hydrolysis

**THE MECHANISM:** The mechanism shown outlines the major stages in carboxypeptidase-catalyzed hydrolysis of a peptide in which the C-terminal amino acid is phenylalanine. Proton transfers accompany stages 2 and 3 but are not shown. Only the major interactions of the substrate with the carboxypeptidase side chains are shown although others may also be involved.

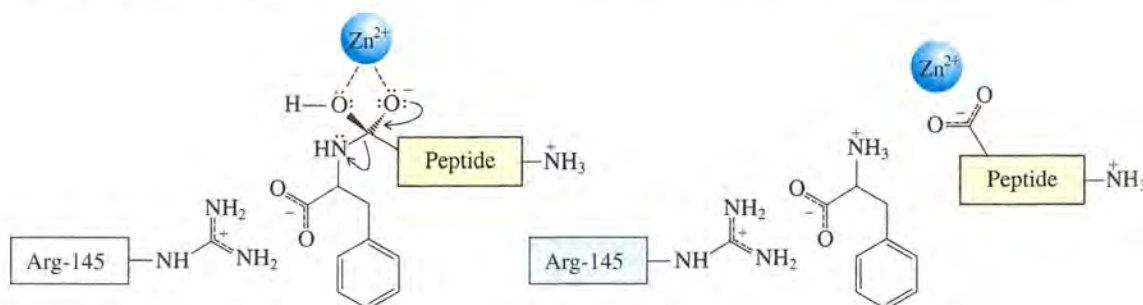
**Stage 1:** The peptide is positioned in the active site by an electrostatic bond between its negatively charged C-terminal carboxylate and a positively charged arginine side chain of the enzyme. Also at the active site,  $\text{Zn}^{2+}$  engages in Lewis acid/Lewis base interactions with His-69 and His-196 and an electrostatic attraction with the negatively charged carboxylate of Glu-72. These ligands are shown here but will be omitted in subsequent steps for simplicity.

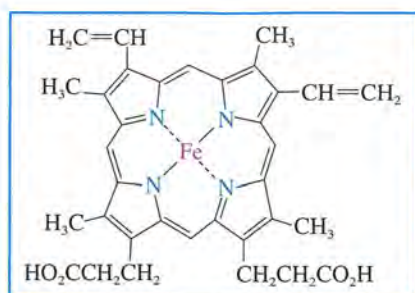


**Stage 2:** Water adds to the carbonyl group of the peptide bond. The rate of this nucleophilic addition is accelerated by coordination of the carbonyl oxygen to  $\text{Zn}^{2+}$  and/or to one of the N—H protons of Arg-127 (not shown). The product is a tetrahedral intermediate stabilized by coordination to zinc. Stabilization of the tetrahedral intermediate may be the major factor for the rapid rate of the carboxypeptidase-catalyzed hydrolysis.

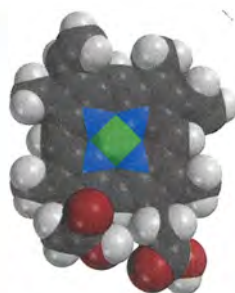


**Stage 3:** The tetrahedral intermediate dissociates to the C-terminal amino acid (phenylalanine in this case). Subsequent steps restore the active site.





(a)



(b)

Figure 25.20

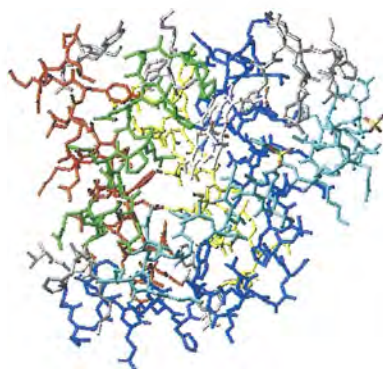
Heme shown as (a) a structural drawing and as (b) a space-filling model. The space-filling model shows the coplanar arrangement of the groups surrounding iron.

## 25.21 Coenzymes

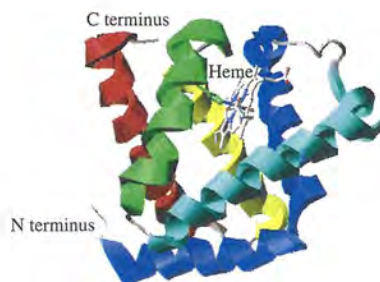
The number of chemical processes that protein side chains can engage in is rather limited. Most prominent among them are proton donation, proton abstraction, and nucleophilic addition to carbonyl groups. In many biological processes a richer variety of reactivity is required, and proteins often act in combination with substances other than proteins to carry out the necessary chemistry. These substances are called **cofactors**, and they can be organic or inorganic and strongly or weakly bound to the enzyme. Among cofactors that are organic molecules, the term **coenzyme** is applied to those that are not covalently bound to the enzyme, and **prosthetic group** to those that are. Acting alone, for example, proteins lack the necessary functionality to be effective oxidizing or reducing agents. They can catalyze biological oxidations and reductions, however, in the presence of a suitable coenzyme. In earlier sections we saw numerous examples of these reactions in which the coenzyme  $\text{NAD}^+$  acted as an oxidizing agent, and others in which  $\text{NADH}$  acted as a reducing agent.

Heme (Figure 25.20) is an important prosthetic group in which iron(II) is coordinated with the four nitrogen atoms of a type of tetracyclic aromatic substance known as a *porphyrin*. The oxygen-storing protein of muscle, myoglobin, represented schematically in Figure 25.21, consists of a heme group surrounded by a protein of 153 amino acids. Four of the six available coordination sites of  $\text{Fe}^{2+}$  are taken up by the nitrogens of the porphyrin, one by a histidine residue of the protein, and the last by a water molecule. Myoglobin stores oxygen obtained from the blood by formation of an  $\text{Fe}-\text{O}_2$  complex. The oxygen displaces water as the sixth ligand on iron and is held there until needed. The protein serves as a container for the heme and prevents oxidation of  $\text{Fe}^{2+}$  to  $\text{Fe}^{3+}$ , an oxidation state in which iron lacks the ability to bind oxygen. Separately, neither heme nor the protein binds oxygen in aqueous solution; together, they do it very well.

Pyridoxal 5'-phosphate (see Section 25.6) is a coenzyme.



(a)



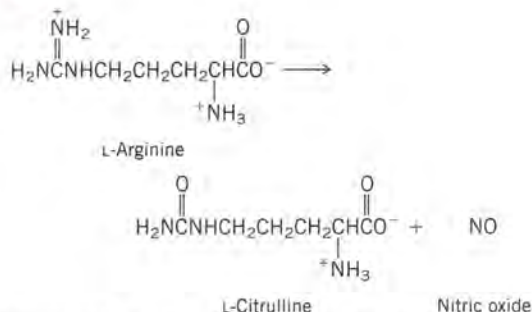
(b)

Figure 25.21

The structure of sperm-whale myoglobin displayed as (a) a tube model and (b) a ribbon diagram. There are five separate regions of  $\alpha$  helix in myoglobin, which are shown in different colors to distinguish them more clearly. The heme portion is included in both drawings, but is easier to locate in the ribbon diagram, as is the histidine side chain that is attached to the iron of heme.

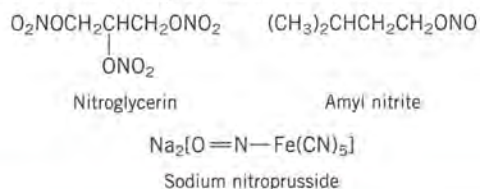
## Oh NO! It's Inorganic!

The amino acid L-arginine undergoes an interesting biochemical conversion.



Our experience conditions us to focus on the organic components of the reaction—L-arginine and L-citrulline—and to give less attention to the inorganic one—nitric oxide (nitrogen monoxide, NO). To do so, however, would lead us to overlook one of the most important discoveries in biology in the last quarter of the twentieth century.

Our story starts with the long-standing use of nitroglycerin to treat the chest pain that characterizes angina, a condition in diseases such as atherosclerosis in which restricted blood flow to the heart muscle itself causes it to receive an insufficient amount of oxygen. Placing a nitroglycerin tablet under the tongue provides rapid relief by expanding the blood vessels feeding the heart. A number of other nitrogen-containing compounds such as amyl nitrite and sodium nitroprusside exert a similar effect.



A chemical basis for their action was proposed in 1977 by Ferid Murad who showed that all were sources of NO, thereby implicating it as the active agent.

Three years later, Robert F. Furchgott discovered that the relaxing of smooth muscles, such as blood vessel walls, was stimulated by an unknown substance produced in the lining of the blood vessels (the *endothelium*). He called this substance the *endothelium-dependent relaxing factor*, or EDRF and, in 1986, showed that EDRF was NO. Louis J. Ignarro reached the same conclusion at about the same time. Further support was provided by Salvador Moncada who showed that endothelial cells did indeed produce NO and that the L-arginine-to-L-citrulline conversion was responsible.

The initial skepticism that greeted the idea that NO, which is (a) a gas, (b) toxic, (c) inorganic, and (d) a free radical, could be a biochemical messenger was quickly overcome. An avalanche of results confirmed not only NO's role in smooth-muscle relaxation, but added more and more examples to an ever-expanding list of NO-stimulated biochemical processes. Digestion is facilitated by the action of NO on intestinal muscles. The drug *Viagra* (*sildenafil citrate*), prescribed to treat erectile dysfunction, works by increasing the concentration of a hormone, the release of which is signaled by NO. A theory that NO is involved in long-term memory receives support from the fact the brain is a rich source of the enzyme *nitric oxide synthase* (NOS), which catalyzes the formation of NO from L-arginine. NO even mediates the glow of fireflies. They glow nonstop when placed in a jar containing NO, but not at all when measures are taken to absorb NO.

Identifying NO as a signaling molecule in biological processes clearly justified a Nobel Prize. The only mystery was who would get it. Nobel Prizes are often shared, but never among more than three persons. Although four scientists—Murad, Furchgott, Ignarro, and Moncada—made important contributions, the Nobel committee followed tradition and recognized only the first three of them with the 1998 Nobel Prize in Physiology or Medicine.

## 25.22 Protein Quaternary Structure: Hemoglobin

Rather than existing as a single polypeptide chain, some proteins are assemblies of two or more chains. The manner in which these subunits are organized is called the **quaternary structure** of the protein.

*Hemoglobin* is the oxygen-carrying protein of blood. It binds oxygen at the lungs and transports it to the muscles, where it is stored by myoglobin. Hemoglobin binds oxygen in very much the same way as myoglobin, using heme as the prosthetic group. Hemoglobin is much larger than myoglobin, however, having a molecular weight of 64,500, whereas that of myoglobin is 17,500; hemoglobin contains four heme units, myoglobin only one. Hemoglobin is an assembly of four hemes and four protein chains, including two identical chains called the *alpha chains* and two identical chains called the *beta chains*.

Some substances, such as CO, form strong bonds to the iron of heme, strong enough to displace O<sub>2</sub> from it. Carbon monoxide binds 30–50 times more effectively than oxygen to myoglobin and hundreds of times better than oxygen to hemoglobin.

Strong binding of CO at the active site interferes with the ability of heme to perform its biological task of transporting and storing oxygen, with potentially lethal results.

How function depends on structure can be seen in the case of the genetic disorder *sickle cell anemia*. This is a debilitating, sometimes fatal, disease in which red blood cells become distorted (“sickle-shaped”) and interfere with the flow of blood through the capillaries. This condition results from the presence of an abnormal hemoglobin in affected people. The primary structures of the beta chain of normal and sickle cell hemoglobin differ by a single amino acid out of 146; sickle cell hemoglobin has valine in place of glutamic acid as the sixth residue from the N terminus of the  $\beta$  chain. A tiny change in amino acid sequence can produce a life-threatening result! This modification is genetically controlled and probably became established in the gene pool because bearers of the trait have an increased resistance to malaria.

## 25.23 G-Coupled Protein Receptors

Biological receptors have been mentioned a few times in this book, in the context of organic molecules that bind to them. In Section 24.2 of the previous chapter, we described anandamide, a lipid that binds to the same receptor in the brain as the cannabinoids. In Section 25.14 of this chapter, we mentioned calcitonin, a peptide that regulates blood calcium levels and is used in the treatment of osteoporosis. The biological receptors for anandamide and for calcitonin both belong to a very large class of protein receptors known as **G-coupled protein receptors**, or GPCRs. The “G” stands for guanine in “guanine nucleotide-binding proteins.” GPCRs occur throughout the body and function as “molecular switches” that regulate many physiological processes.

GPCRs span the cell membrane (see Figure 24.4). When GPCRs bind their specific ligand, such as a lipid, peptide, or ion, they undergo a conformational change, which results in the transduction of a signal across the membrane. The details of how signaling occurs are not completely understood. The conformational change may result in an interaction with a nearby G protein, which in turn can activate enzymes or ion channels (Figure 25.22).

In the biochemistry of vision (see Chapter 17), the interaction of rhodopsin, a GPCR, with light results in the photo-induced isomerization of cis retinal imine to trans retinal imine. This causes conformational changes in rhodopsin, which ultimately result in the *closing* of an ion channel, polarization of the cell membrane, and a nerve impulse that is transmitted to the brain in vision.

G-coupled protein receptors are involved in many diseases. It is estimated that nearly one-half of prescription drugs target GPCRs, in the treatment of cancer, cardiac malfunction, inflammation, pain, and disorders of the central nervous system.

Signal transduction pathways in biochemistry involve a relay of molecular interactions that ultimately produce an intracellular response. These pathways can serve as a connection between events at the cell surface and gene expression in the nucleus.

GDP and GTP are abbreviations for the nucleotides guanosine 5'-diphosphate and guanosine 5'-triphosphate. Nucleotides are discussed in Section 26.2.

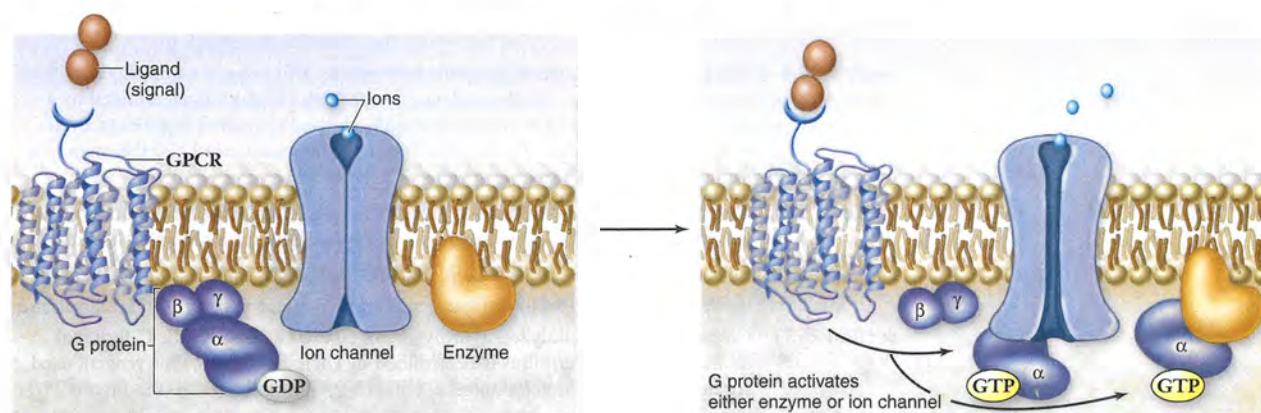


**Figure 25.22**

Signal transduction is initiated by the binding of a G-coupled protein receptor (GPCR) to the ligand on the exterior of the cell. Interactions with a nearby G protein inside the cell results in the exchange of bound GDP to GTP and in the release of one of its subunits in a GTP-bound form that activates an ion channel or enzyme.

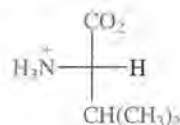
## 25.24 SUMMARY

This chapter revolves around **proteins**. The first half describes the building blocks of proteins, progressing through **amino acids** and **peptides**. The second half deals with proteins themselves.

**Section 25.1** The 20 amino acids listed in Table 25.1 are the building blocks of proteins. All are  $\alpha$ -amino acids.

**Section 25.2** Except for glycine, which is achiral, all of the  $\alpha$ -amino acids in Table 25.1 are chiral and have the L configuration at the  $\alpha$  carbon.

**Section 25.3** The most stable structure of a neutral amino acid is a **zwitterion**. The pH of an aqueous solution at which the concentration of the zwitterion is a maximum is called the isoelectric point (pI).



Fischer projection of  
L-valine in its zwitterionic form

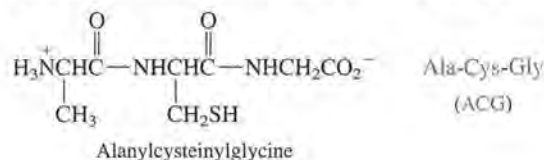
**Section 25.4** Amino acids are synthesized in the laboratory from

1.  $\alpha$ -Halo acids by reaction with ammonia
2. Aldehydes by reaction with ammonia and cyanide ion (the Strecker synthesis)
3. Alkyl halides by reaction with the enolate anion derived from diethyl acetamidomalonate

**Section 25.5** Amino acids undergo reactions characteristic of the amino group (e.g., amide formation) and the carboxyl group (e.g., esterification). Amino acid side chains undergo reactions characteristic of the functional groups they contain.

**Section 25.6** Among the biochemical reactions of  $\alpha$ -amino acids, several use pyridoxal 5'-phosphate as a coenzyme. These reactions involve bonds to the  $\alpha$  carbon and include transamination, decarboxylation, and racemization.

**Section 25.7** An amide linkage between two  $\alpha$ -amino acids is called a **peptide bond**. By convention, peptides are named and written beginning at the N terminus.



**Section 25.8** The **primary structure** of a peptide is its amino acid sequence plus any disulfide bonds between two cysteine residues. The primary structure is determined by a systematic approach in which the protein is cleaved to smaller fragments, even individual amino acids. The smaller fragments are sequenced and the main sequence deduced by finding regions of overlap among the smaller peptides.

**Section 25.9** Complete hydrolysis of a peptide gives a mixture of amino acids. An amino acid analyzer identifies the individual amino acids and determines their molar ratios.

**Section 25.10** Selective hydrolysis can be accomplished by using enzymes to catalyze cleavage at specific peptide bonds.

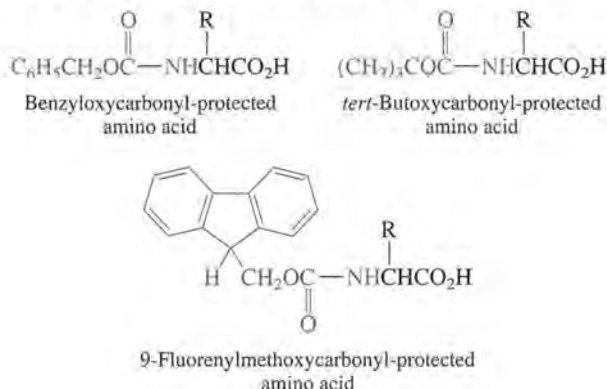
**Section 25.11** Carboxypeptidase-catalyzed hydrolysis can be used to identify the C-terminal amino acid. The N terminus is determined by chemical means. One reagent used for this purpose is Sanger's reagent, 1-fluoro-2,4-dinitrobenzene (see Figure 25.7).

**Section 25.12** The procedure described in Sections 25.8–25.11 was used to determine the amino acid sequence of insulin.

**Section 25.13** Modern methods of peptide sequencing follow a strategy similar to that used to sequence insulin, but are automated and can be carried out on a small scale. A key feature is repetitive N-terminal amino acid identification using the **Edman degradation**.

**Section 25.14** Synthesis of a peptide of prescribed sequence requires the use of protecting groups to minimize the number of possible reactions.

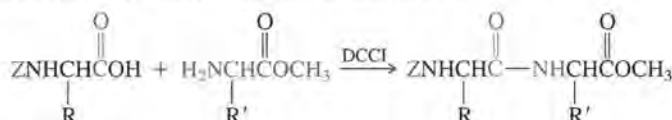
**Section 25.15** Amino-protecting groups include *benzyloxycarbonyl* (Z), *tert-butoxycarbonyl* (Boc), and *9-fluorenylmethoxycarbonyl* (Fmoc).



Hydrogen bromide may be used to remove either the benzyloxycarbonyl or *tert*-butoxycarbonyl protecting group. The benzyloxycarbonyl protecting group may also be removed by catalytic hydrogenolysis. Fmoc is removed in base.

**Section 25.16** Carboxyl groups are normally protected as benzyl, methyl, or ethyl esters. Hydrolysis in dilute base is normally used to deprotect methyl and ethyl esters. Benzyl protecting groups are removed by hydrogenolysis.

**Section 25.17** Peptide bond formation between a protected amino acid having a free carboxyl group and a protected amino acid having a free amino group can be accomplished with the aid of *N,N'*-dicyclohexylcarbodiimide (DCCI).



**Section 25.18** In the Merrifield method the carboxyl group of an amino acid is anchored to a solid support and the chain extended one amino acid at a time. When all the amino acid residues have been added, the polypeptide is removed from the solid support.

**Section 25.19** Two **secondary structures** of proteins are particularly prominent. The *pleated  $\beta$  sheet* is stabilized by hydrogen bonds between N—H and C=O groups of adjacent chains. The  *$\alpha$  helix* is stabilized by hydrogen bonds within a single polypeptide chain.

**Section 25.20** The folding of a peptide chain is its **tertiary structure**. The tertiary structure has a tremendous influence on the properties of the peptide and the biological role it plays. The tertiary structure is normally determined by X-ray crystallography.

Many globular proteins are enzymes. They accelerate the rates of chemical reactions in biological systems, but the kinds of reactions that take place are the fundamental reactions of organic chemistry. One way in which enzymes accelerate these reactions is by bringing reactive functions together in the presence of catalytically active functions of the protein.

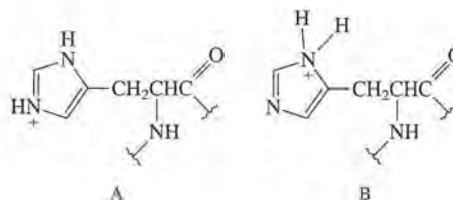
**Section 25.21** Often the catalytically active functions of an enzyme are nothing more than proton donors and proton acceptors. In many cases a protein acts in cooperation with a **coenzyme**, a small molecule having the proper functionality to carry out a chemical change not otherwise available to the protein itself.

**Section 25.22** Many proteins consist of two or more chains, and the way in which the various units are assembled in the native state of the protein is called its **quaternary structure**.

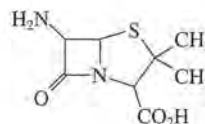
**Section 25.23** G-coupled protein receptors are transmembrane proteins that function as molecular switches in many physiological processes.

## PROBLEMS

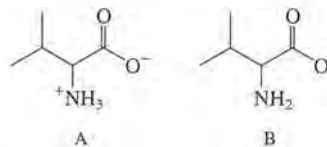
- 25.28 The imidazole ring of the histidine side chain acts as a proton acceptor in certain enzyme-catalyzed reactions. Which is the more stable protonated form of the histidine residue, A or B? Why?



- 25.29 Which two  $\alpha$ -amino acids are the biosynthetic precursors of the penicillins?



- 25.30 (a) Use the data in Table 25.2 and the Henderson–Hasselbalch equation to calculate the ratio  $\frac{[A]}{[B]}$  at pH = 7.

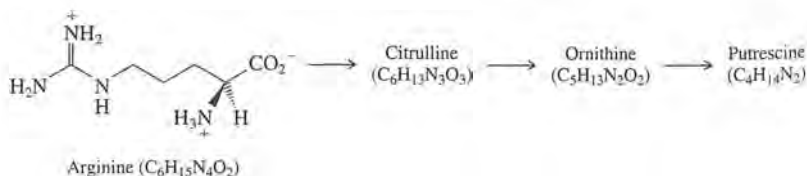


- (b) At what pH is [A] the largest?

- 25.31  $\alpha$ -Amino acids are not the only compounds that exist as zwitterions. *p*-Aminobenzenesulfonic acid (sulfanilic acid) is normally written in the form shown but its zwitterionic form is more stable. Write a structural formula for the zwitterion.

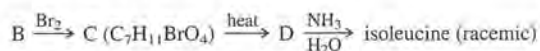
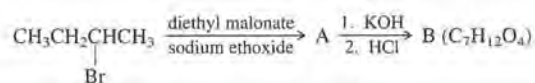


- 25.32 Putrescine, citrulline, and ornithine are products of arginine metabolism. The molecular formulas are given for the form in which each exists at pH = 7. The net charge corresponding to each formula is zero, +1, and +2. Suggest a reasonable structure for each species.



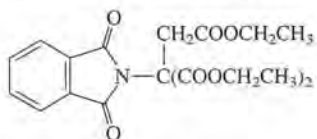
- 25.33 Acrylonitrile ( $\text{H}_2\text{C}=\text{CHC}\equiv\text{N}$ ) readily undergoes conjugate addition when treated with nucleophilic reagents. Describe a synthesis of  $\beta$ -alanine ( $\text{H}_3\text{N}^+\text{CH}_2\text{CH}_2\text{CO}_2^-$ ) that takes advantage of this fact.

- 25.34 (a) Isoleucine has been prepared by the following sequence of reactions. Give the structure of compounds A through D isolated as intermediates in this synthesis.

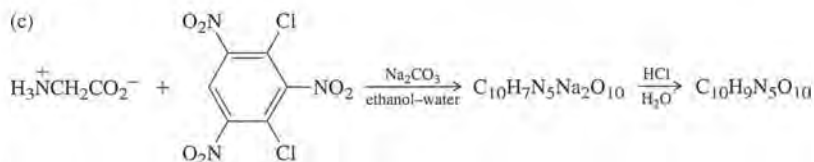
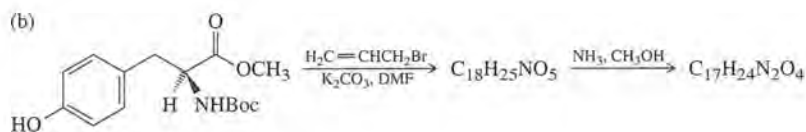
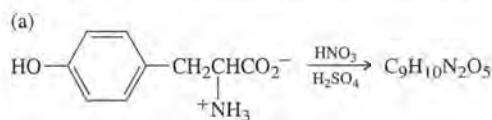


(b) An analogous procedure has been used to prepare phenylalanine. What alkyl halide would you choose as the starting material for this synthesis?

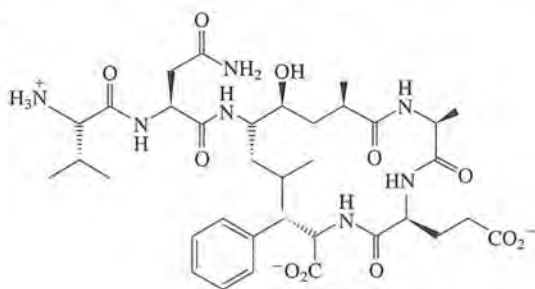
- 25.35 Hydrolysis of the following compound in concentrated hydrochloric acid for several hours at 100°C gives one of the amino acids in Table 25.1. Which one? Is it optically active?



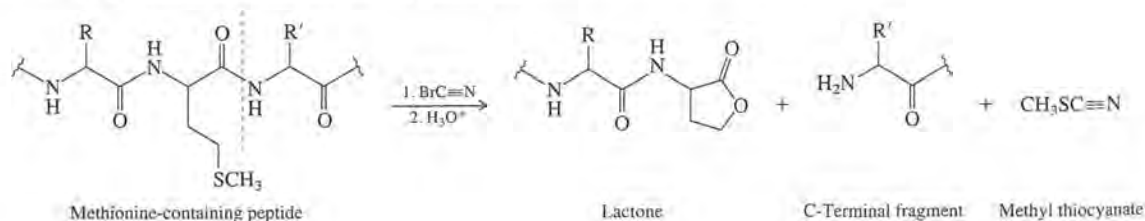
- 25.36 Identify the major product in each of the following reactions involving amino acids.



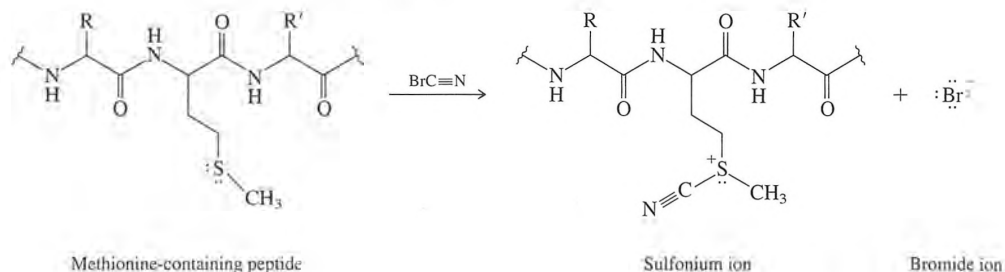
- 25.37 The synthetic peptide shown here is an inhibitor of the enzyme  $\beta$ -secretase, which plays a role in the development of Alzheimer's disease. It contains five amino acids from Table 25.1. Which ones?



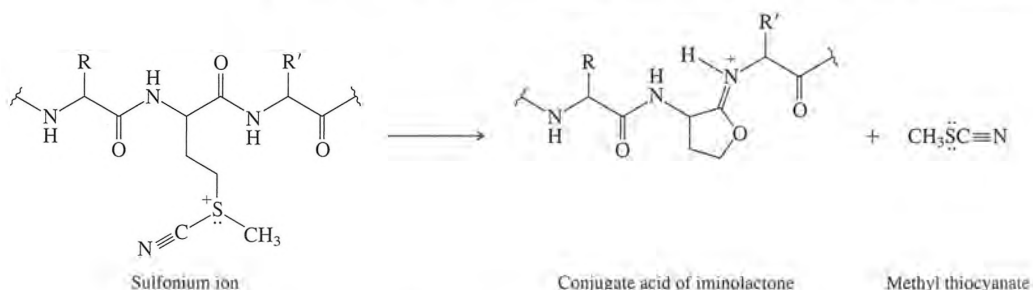
- 25.38 Cyanogen bromide ( $\text{BrC}\equiv\text{N}$ ) cleaves peptides specifically between the carbonyl group of a methionine residue and the next amino acid.



- (a) The mechanism of this process begins with the reaction of cyanogen bromide with the peptide to give a sulfonium ion:



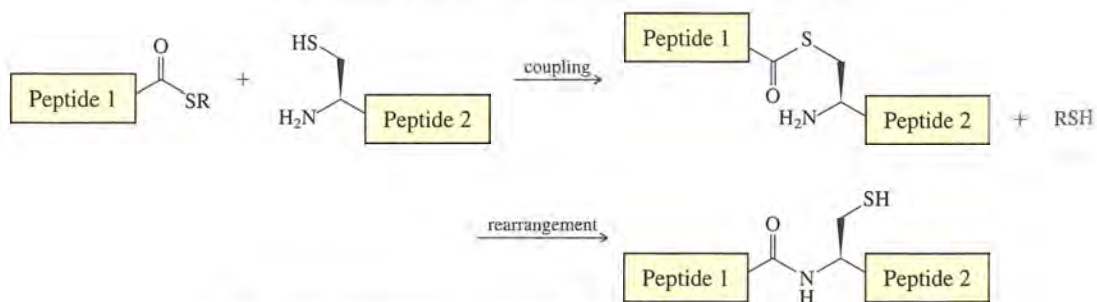
The sulfonium ion then cyclizes to an iminolactone with loss of methyl thiocyanate.



Using curved arrows to show the flow of electrons, suggest a reasonable mechanism for this cyclization.

- (b) Hydrolysis of the conjugate acid of the iminolactone to the corresponding lactone cleaves the peptide chain. Write a stepwise mechanism for this reaction.

**25.39** *Native chemical ligation (NCL)* is a method for coupling peptide chains. It requires that one of the peptides has an N-terminal cysteine, and the C terminus of the other peptide be a thioester. The thiol of the N-terminal cysteine of peptide 2 reacts with the thioester of peptide 1 to give an initial product that contains a “nonnative” thioester linkage. The initial product rearranges to a more stable, amide-linked (native) peptide. Write mechanisms for the coupling and rearrangement steps in NCL.



- 25.40** If you synthesized the tripeptide Leu-Phe-Ser from amino acids prepared by the Strecker synthesis, how many stereoisomers would you expect to be formed?
- 25.41** Automated amino acid analysis of peptides containing asparagine (Asn) and glutamine (Gln) residues gives a peak corresponding to ammonia. Why?
- 25.42** What are the products of each of the following reactions? Your answer should account for all the amino acid residues in the starting peptides.
- (a) Reaction of Leu-Gly-Ser with 1-fluoro-2,4-dinitrobenzene
- (b) Hydrolysis of the compound in part (a) in concentrated hydrochloric acid (100°C)

- (c) Treatment of Ile-Glu-Phe with  $\text{C}_6\text{H}_5\text{N}=\text{C}=\text{S}$ , followed by hydrogen bromide in nitromethane
- (d) Reaction of Asn-Ser-Ala with benzyloxycarbonyl chloride
- (e) Reaction of the product of part (d) with *p*-nitrophenol and *N,N'*-dicyclohexylcarbodiimide
- (f) Reaction of the product of part (e) with the ethyl ester of valine
- (g) Hydrogenolysis of the product of part (f) by reaction with  $\text{H}_2$  over palladium
- 25.43 The first 32 amino acids from the N terminus of the protein *bovine angiogenin* were determined by Edman degradation and have the sequence:

AQDDYRIHFLTQH YDAKPKGRNDEYCFNMMK

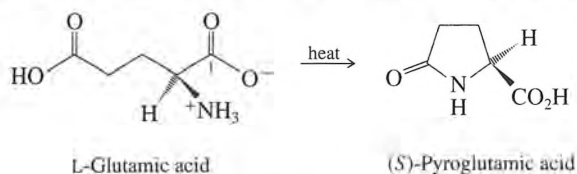
- (a) Identify the sites of cleavage during trypsin-catalyzed hydrolysis of this protein.
- (b) What are the cleavage sites using chymotrypsin?
- 25.44 *Somatostatin* is a tetradecapeptide of the hypothalamus that inhibits the release of pituitary growth hormone. Its amino acid sequence has been determined by a combination of Edman degradations and enzymic hydrolysis experiments. On the basis of the following data, deduce the primary structure of somatostatin:
1. Edman degradation gave PTH-Ala.
  2. Selective hydrolysis gave peptides having the following indicated sequences:  
 Phe-Trp  
 Thr-Ser-Cys  
 Lys-Thr-Phe  
 Thr-Phe-Thr-Ser-Cys  
 Asn-Phe-Phe-Trp-Lys  
 Ala-Gly-Cys-Lys-Asn-Phe
  3. Somatostatin has a disulfide bridge.
- 25.45 What protected amino acid would you anchor to the solid support in the first step of a synthesis of oxytocin (see Figure 25.6) by the Merrifield method?

## Descriptive Passage and Interpretive Problems 25

### Amino Acids in Enantioselective Synthesis

Organic chemists speak of a "chiral pool," which comprises those naturally occurring compounds that are readily available as a single enantiomer and capable of being used as starting materials for the enantioselective synthesis of other chiral molecules. Amino acids are well represented in the chiral pool. All except glycine have at least one chirality center and, although L-amino acids are more abundant and less expensive than their D-enantiomers, both are available.

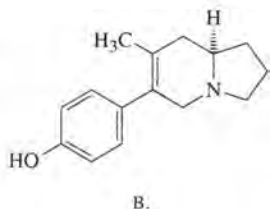
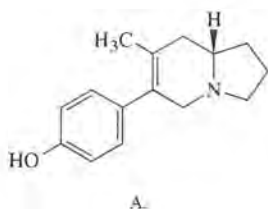
Most of the standard amino acids have served as starting materials for enantioselective syntheses. One of the most widely used is L-glutamic acid and its lactam (*S*)-pyroglutamic acid, which is easily prepared by heating an aqueous solution of L-glutamic acid in a sealed container.



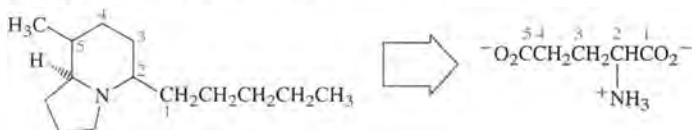
With three functional groups in a compound with only five carbons, L-glutamic acid and (*S*)-pyroglutamic acid provide access to more complex molecules via functional-group manipulation.

Many of the syntheses are lengthy, and most contain some specialized reactions. The following problems emphasize the planning aspect of amino acid-based enantioselective syntheses starting with either L-glutamic acid or (*S*)-pyroglutamic acid. The few reactions that are included are either familiar or similar to those covered in earlier chapters.

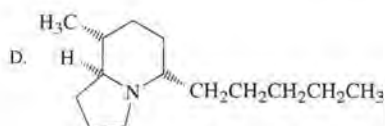
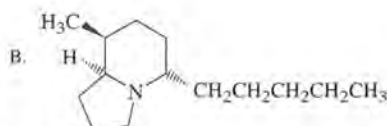
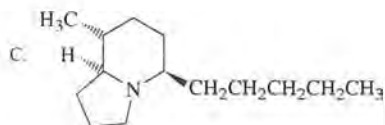
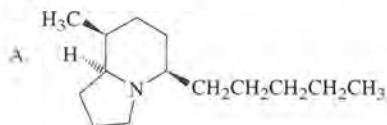
- 25.46 (S)-Pyroglutamic acid was used as the starting material in a synthesis of (+)-ipalbidine, an analgesic alkaloid obtained from the seeds of the white moonflower, *Ipomoea alba*. No bonds are made or broken to the chirality center of (S)-pyroglutamic acid in this synthesis. Which of the following is the structure of (+)-ipalbidine?



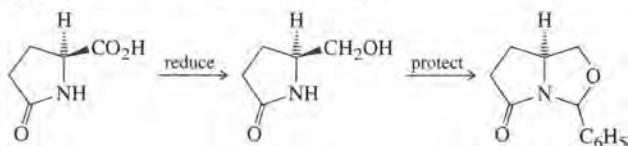
- 25.47 A synthesis of poison-dart frog toxin has been described that begins with L-glutamic acid.



The numbered carbons in the product correspond to the same numbered carbons in L-glutamic acid. No bonds to carbon-2 are made or broken in the synthesis. If the configuration at carbon-5 in the product is *R*, which of the following best represents the stereochemistry of the frog toxin?



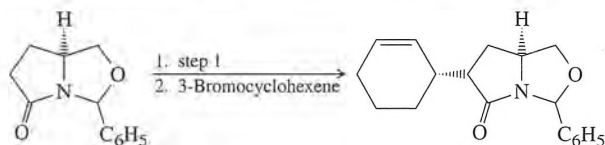
- 25.48 One synthesis of fosinopril, a drug used to combat high blood pressure, starts with the reduction of (S)-pyroglutamic acid to the primary alcohol, followed by protection of the OH and NH groups.



What reaction conditions are appropriate for the protection step?

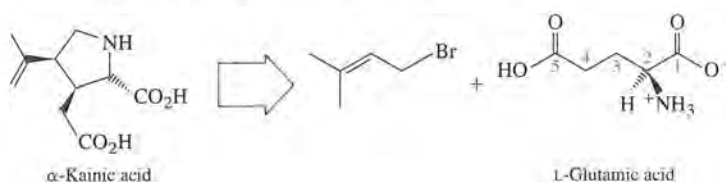
- A.  $\text{C}_6\text{H}_5\text{CCl}$ , pyridine  
 B.  $\text{C}_6\text{H}_5\text{CH}=\text{O}$ , *p*-toluenesulfonic acid, toluene, heat  
 C.  $\text{C}_6\text{H}_5\text{CO}_2\text{CH}_3$   
 D.  $\text{C}_6\text{H}_5\text{CH}_2\text{Br}$ ,  $\text{Na}_2\text{CO}_3$ , acetone

- 25.49 The *N,O*-protected compound formed in the preceding problem was alkylated with 3-bromocyclohexene.



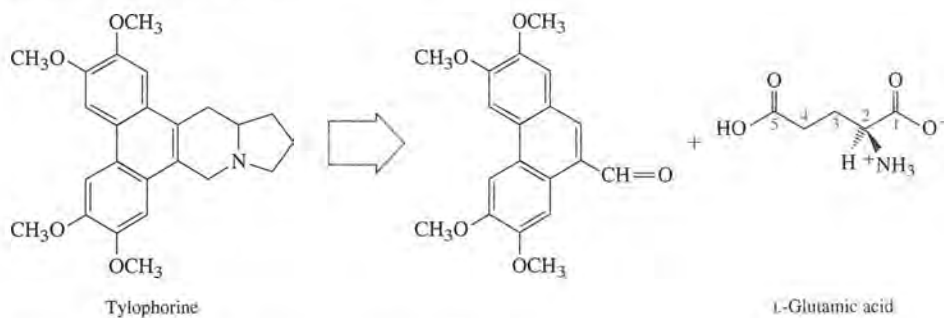
What reaction conditions are appropriate for step 1?

- A.  $\text{LiAlH}_4$ , diethyl ether  
 B.  $\text{NaOCH}_2\text{CH}_3$ , ethanol  
 C. Mg, diethyl ether  
 D.  $[(\text{CH}_3)_2\text{CH}]_2\text{NLi}$ , tetrahydrofuran
- 25.50  $\alpha$ -Kainic acid is a neurotoxin produced by certain algae. Several enantioselective syntheses of it have been described, one of which is based on 1-bromo-3-methyl-2-butene and L-glutamic acid.



Which carbon of L-glutamic acid is involved in the only carbon-carbon bond forming step of the synthesis?

- A. C-1  
 B. C-2  
 C. C-3  
 D. C-4  
 E. C-5
- 25.51 (*S*)-Tylophorine is an alkaloid isolated from a plant that grows in India and Southeast Asia, which is of interest as a potential antitumor drug. It has been synthesized by a multistep procedure based on L-glutamic acid and the phenanthrene derivative shown.



Which carbon of L-glutamic acid is involved in the only carbon-carbon bond forming step of the synthesis?

- A. C-1  
 B. C-2  
 C. C-3  
 D. C-4  
 E. C-5